نماذج إحصائية خطية تطبيقية

انحدار ، تحليل تباين وتصاميم تجريبية

الجزء الأول (الانحدار)

تأليف

جون نتر

ويليام وازرمان

ميخائيل كتنر

ترجمة

أ.د. أنيس إسماعيل كنجو   أ.د. عبد الحميد بن عبد الله الزيد

د. إبراهيم بن عبد العزيز الواصل   د.الحسيني عبد البر راضي



النشر العلمي و المطابع

جامعة الملك سعود

اهداءات ٢٠٠٤


الله
بسم الله الرحمن الرحيم



نماذج إحصائية خطية تطبيقية

انحدار، تحليل تباين وتصاميم تجريبية

الجزء الأول (الانحدار)

تأليـف

جون نتر — جامعة جورجيا

ويليام وازرمان — جامعة سيراكاس

ميخائيل كتنر — جامعة إيموري

ترجـمـة

أ. د. أليس اسماعيل كنجو

أ.د. عبدالحميد بن عبدالله الزيد

د. ابراهيم بن عبدالعزيز الواصل

د. الحسيني عبدالبر راضي

قسم الإحصاء وبحوث العمليات - كلية العلوم - جامعة الملك سعود



النشر العلمي والمطابع - جامعة الملك سعود

ص.ب ٦٨٩٥٣ - الرياض ١١٥٣٧ - المملكة العربية السعودية

هذه ترجمة عربية مصرح بها لكتاب:

Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs (Third Edition)
By: John Neter, William Wasserman & Michael Kutner

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

نتر، جون

نماذج إحصائية خطية تطبيقية: انحدار، تحليل تباين وتصاميم تجريبية: الجزء الأول (الانحدار)/جون نتر، ويليام وازرمان وميخائيل كتنر؛ ترجمة أنيس اسماعيل كنجو [وآخرون]-الرياض

٧٥٥ص: ١٧سم × ٢٥سم

ردمك ٩-١٣٧-٣٧-٩٩٦٠ (مجموعة)

١-٩٤٢-٠٥-٩٩٦٠ (ج١)

(الجزء الأول: الانحدار)

١-الجبر الخطي ٢-المعادلات الخطية ٣-الاحصاء الرياضي

أ-وازرمان، وليام (م. مشارك) ب-كتنر، ميخائيل (م. مشارك) ج-كنجو، أنيس اسماعيل (مترجم)

د-العنوان

ديوي ٥١٢,٥ ٢١/١٠٦٥

رقم الإيداع: ٢١/١٠٦٥

حكمت هذا الكتاب لجنة متخصصة شكلها المجلس العلمي بالجامعة، وقد وافق المجلس العلمي على نشره في اجتماعه الثالث عشر للعام الدراسي ١٤١٦/١٤١٧هـ المعقود بتاريخ ١٤١٦/٨/٩هـ الموافق ١٩٩٥/١٢/٣١م.

جامعة الملك سعود ١٤٢١هـ

مقدمة المترجمين

الحمد لله وحده والصلاة والسلام على من لا نبيّ بعده سيدنا محمد بن عبد الله الهادي الأمين والمرسل بلسان عربي مبين، وبعد فقد وقع اختيارنا على كتاب نماذج إحصائية خطية تطبيقية لأسباب عدة نوجزها فيما يلي:

١ - يتطرق الكتاب لتشكيلة واسعة من التطبيقات الإحصائية تتناول بصورة شاملة تقريبا تحليل الانحدار وأهم ما يحتاجه الباحث والدارس من تطبيقات تحليل التباين وتصميم التجارب، ويعرج في هذه الرحلة الطويلة في دنيا الطرائقية الإحصائية على بعض من تطبيقات السلاسل الزمنية والإحصاء اللامعلمي.

٢ - يتميز الكتاب بعرض واضح وميسر لأساسيات الطرق الإحصائية، وللمفاهيم الرئيسية التي تشكل خلفيتها النظرية، مما يرفع بشكل ملحوظ من قدرة الدار على التطبيق السليم، وتجنب الشطط والاستخدام المضلل للإحصاء، ويعينه على فهم النتائج التي يحصل عليها، وتفسيرها تفسيرا صحيحا، وعرضها بدقة وإحكام، وكان ذلك ثمرة تعاون ثلاثة مؤلفين ممن برعوا في مجال الإحصاء التطبيقي بالإضافة إلى خبرة عدد وافر من المراجعين، وحصيلة سنوات طويلة من الخبرة الميدانية الواسعة.

٣ - يتميز الكتاب بتشكيلة فريدة من المسائل الميدانية المأخوذة، من تطبيقات واقعية في مجالات شتى، شملت العلوم الاجتماعية والأحيائية وعلوم الإدارة والاقتصاد والصناعة وغيرها، وهو بما يحتويه من الأمثلة والمسائل والتمارين والمشاريع والبيانات الإحصائية الواقعية، يشكل من حيث الكم والكيف مرجعا لا غنى عنه لقاعدة واسعة من الباحثين والدارسين والمستفيدين.

٤ - وإلى جانب شمولية العرض يتميز الكتاب بحداثة العرض، وإذا خرجت آخر طبعة للكتاب، وهي الطبعة الثالثة، في التسعينات فقد احتوت عددا من التقنيات الحديثة التي ظهرت للمرة الأولى في السبعينات والثمانينات، لاسيما في مجال التشخيص لعلّة أو علل

الإحصائي المستخدم لتحليل البيان، وآفاق الاستفادة منه في مجالات التقدير أو التنبؤ أو السيطرة. وكان لا بد للكتاب، وقد ارتدى ثوب الحداثة هذا، أن يعتمد بقوة على استخدام الحاسب الآلي، ويتجنب الغوص في صيغ الحسابات اليدوية التقليدية التي تستهلك جزءا غير قليل من الكتب الطرائقية التقليدية.

٥ ـ نجح الكتاب في عرض ثلاثة مواضيع متفرقة هي تحليل الانحدار، وتحليل التباين، وتحليل التجارب المصممة، في إطار موحّد هو إطار النماذج الخطية التطبيقية، مما يسمح إضافة إلى الفوائد النظرية، بالاستفادة من أفضل ما تضمنته الحزم الإحصائية الحديثة، ويمكّن من الاستخدام الأمثل للحاسوب في التحليل الإحصائي.

ونظرا لوفرة المواد التي يقدمها الكتاب فقد تقرّر بعد موافقة الناشر إصدار الترجمة في جزئين، يتضمن الجزء الأول الفصول الثلاثة عشر الأولى وهي تشمل تحليل الانحدار، ويتضمن الجزء الثاني الفصول الستة عشر الباقية وهي في تحليل التباين وتصميم التجارب وتحليلها، وكانت مساهمات المترجمين أحد عشر فصلا للدكتور أنيس كنجو وستة فصول لكل من الدكتور عبد الحميد الزيد والدكتور إبراهيم الواصل والدكتور الحسيني راضي، كما قام الدكتور أنيس كنجو بمهمة المراجعة العلمية واللغوية للكتاب مما أضفى على أسلوب العرض وحدة لا تخفى، وأدى إلى انسجام العبارة عبر الكتاب بأكمله.

وكانت مسألة المصطلح العلمي تحديا نرجو أن نكون قد وُفّقنا في مواجهته، خاصة وأن العديد من المصطلحات يظهر في العربية، في حدود معلوماتنا، للمرّة الأولى، وبالطبع نرحب بأية مقترحات يتفضل بها الزملاء والقراء سواء تناولت مصطلحا أو تعبيرا.

وكما أشارت مقدمة المؤلفين فقد صُممت الطبعة الثالثة بحيث تغطي مقررات من مستوى المرحلة الجامعية الأولى ومن مستوى الدراسات العليا. فضلا عن استخدامه كمرجع لباحثين في ميادين الإدارة والاقتصاد والعلوم الاجتماعية والصحية والأحيائية. وأملنا كبير في أن يسدّ هذا الكتاب بجزأيه ثغرة في المكتبة الإحصائية العربية، وما أحوجنا إلى سد الثغرات في المكتبة العلمية العربية بجميع فروعها وأجنحتها وليس في الإحصائية منها فقط. فالعربية لغتنا

الجميلة هي كما يصفها المرحوم الأستاذ الدكتور "محمد المبارك" : "غنية من حيث الأبنية والصيغ غنى لا تضارعها فيه لغة أخرى من اللغات الراقية التي تفي بحاجات الإنسان في مثل هذا العصر الذي نحن فيه، وتدل مفردات اللغة العربية دلالة قاطعة على أن العرب صنفوا الوجود تصنيفا شاملا دقيقا منطقيا يدعو إلى الدهشة والتعجب ويدل على مستوى فكري قلّما وصلت إليه الأمم في مثل ذلك الطّور البكر من تاريخ حياتها".

إن المتأمل من الأساتذة والمفكرين العرب في مردود التعليم الجامعي في بلادنا العربية ترتد إليه تأملاته بوافر من الحسرة والألم وشعور قد يصل حد الإحباط. وهو فوق هذا وكرجل استوعب واقع العصر واستشعر آفاق التقدم الحضاري ووتيرته ينظر إلى قومه بين الأقوام التي انتظمها ركب الحضارة المعاصر فيفتقدهم، ويجيل الطرف من حوله يستشف ساعة الفجر فيجدها، ضمن واقعنا العلمي السائد، بعيدة المنال. لا بل يجد الهوة الكبيرة بينه وبين نظيره في العالم المتقدم علميا تزداد اتساعا وعمقا كل يوم وكل ساعة.

إن بناء المكتبة العلمية العربية واجب على كل مستطيع، فما الذي يمنعنا عن إغناء العربية لتصبح لغة علم تذخر بالمصطلح من كل صنف، وتتميز مكتباتها بلجب من المراجع العلمية المعدة بلغة الضاد؟ ثم كيف يمكن لنا تلمّس الطريق إلى هذا الهدف إذا بقي التعليم الجامعي بلغة أجنبية؟ هل نكتب ونترجم لتوضع جهودنا على الرفوف، أم ليتخذها جمهور الطلبة سبيلا ميسرا إلى المعرفة؟ إن ثوب العيرة الذي نرتديه لا يؤهلنا لأكثر من أدوار التمثيل، فالملكات المبدعة تنمو في حضن العربية، ولا يمكن لها أن تزدهر إلا في حماها، ولن ننطلق في بناء مستقبلنا الحضاري ونأمل في استعادة موقع حضاري يليق بتراثنا المرموق إلا عندما تتيسر المعرفة لكل عربي بلغته الأم.

ولما كانت الأعمال بالنيات، وكان لكل امرئ ما نوى، وكانت نوايانا، فيما اخترناه وفيما بذلناه من جهود، خدمة لغة القرآن المجيد وتقديم زاد علمي مفيد، لكل قارئ بالعربية، فالله سبحانه وتعالى نسأل أن يتقبل منا هذه الترجمة عملا صالحا، فهو من وراء القصد، وهو الهادي إلى سواء السبيل.

المترجمون

مقدمة المؤلفين

تُستخدم النماذج الإحصائية الخطية الخاصة بالانحدار، تحليل التباين، والتصاميم التجريبية، اليوم استخداما واسعا في إدارة الأعمال، الاقتصاد، والعلوم الاجتماعية، الصحية والأحيائية. وتحتاج التطبيقات الناجحة لهذه النماذج إلى فهم سليم لكل من الخلفية النظرية والمسائل العملية التي نواجهها عند استخدام النماذج في حالات من واقع الحياة. وبينما تشكل الطبعة الثالثة من نماذج إحصائية خطية تطبيقية، في الأساس، كتابا تطبيقيا، إلا أنها تهدف إلى خلط النظري والتطبيقات بصورة فعّالة، متجنبة الشطط سواء في تقديم النظري بصورة منعزلة أو في طرح عناصر من التطبيقات دون الحاجة إلى فهم أسسها النظرية.

وتختلف الطبعة الثالثة عن الطبعة الثانية في عدد من النواحي المهمة.

١- أضفنا فصلا جديدا في تصاميم القياسات المكررة نظرا لأهميتها الكبرى في العلوم السلوكية وعلوم الحياة. وبالنسبة للقارىء، يشكّل الفصل الثامن والعشرون المضاف مدخلا إلى تصاميم القياسات المكررة مع المتابعة في تصاميم القياسات المكررة ذات العامل الواحد، وفي التصاميم ذات العاملين مع قياسات مكررة لأحد العاملين أو لهما معا، وفي تصاميم الوحدة المنقسمة.

وبالإضافة إلى ذلك فإن الفصل الثاني عشر حول بناء نموذج انحدار قد أعيدت صياغته إلى حد كبير واتّسع كثيرا. ونطور، في هذا الفصل، بالتفصيل عملية بناء نموذج بحيث يستوعب العديد من عناصر هذه العملية، التي نوقشت في فصول سابقة. ونتعرض أيضا لمعالجة موسّعة جدا للتحقق من نماذج انحدار.

٢- توسعنا كثيرا في مناقشة تشخيصات تحليل الانحدار وتحليل التباين وذلك عبر الكتاب بأكمله. ففي ميدان تحليل الانحدار نتابع الآن، من بين التدابير التشخيصية المدروسة، تدابير PRESS; DFFITS; DFBETAS، كما أضفنا أيضا رسومات الانحدار الجزئي، كما ندرس تحويل بوكس - كوكس كتدبير علاجي.

وقد ازددنا أيضا من التأكيد على التشخيصات في تحليل التباين وتصميم التجارب، إذ نقدم عددا أكبر بكثير من الرسومات التشخيصية، كما أضفنا مناقشة رسوم آحتمال طبيعي للتأثيرات الرئيسة المقدَّرة للعوامل.

٣- وقد توسعنا في عدد من المواضيع وأعدنا تنظيمها. ففي ميدان تحليل الانحدار وُحِّدت الآن مناقشة المربعات الدنيا المرجحة ودُرست في سياق الانحدار المتعدد. وقد أعيد تنظيم مناقشة نماذج الانحدار المعيارية، كما دُعم عرض كل من مجاميع المربعات الإضافية والخطية المتعددة من خلال إعادة تنظيم شاملة لها، كما توسعنا في الفصل الثالث عشر وهو فصل الارتباط الذاتي بأن درسنا طريقة هيلدرت – لو (Hildreth - Lu) في تقدير معلمة الارتباط الذاتي، وأضفنا فقرة تتعلق بفترات تنبؤ عند التنبؤ بمشاهدة جديدة. وأضفنا أيضا مناقشة موجزة لطرائقية سطح الاستجابة في الفصل التاسع المتعلق بانحدار كثيرات الحدود.

وفي ميدان تحليل التباين والتصاميم التجريبية، توسّعنا كثيرا في شرح نماذج التحاين، خاصة ماتعلق منها بنماذج التأثيرات العشوائية المختلطة لتصاميم القطاع العشوائي، التصاميم الحاضنة، تصاميم القياسات المكررة، وتصاميم المربع اللاتيني. وعلى وجه الخصوص أكّدنا على التقابل بين نموذج تحاين والبنية الارتباطية للمشاهدات.

وبالإضافة إلى ذلك، فقد عززنا مناقشة مفهوم القوة وتخطيط حجوم العينات من منظور العلاقات الوثيقة بين هذين الموضوعين.

وقد اتسعت أيضا مناقشة التحاين متعدد العوامل وذلك عندما لاتكون متوسطات المعالجات متساوية الأهمية.

٤- وقد عززنا، عبر الكتاب، التكامل بين التصاميم التجريبية ودراسات المشاهدة، مبتدئين بمناقشة الحصول على بيانات لتحليل الانحدار في الفصل الثاني.

٥- وقمنا، عبر الكتاب، بتنقيح شامل في العرض مستندين الى الخبرة الميدانية ضمن الفصل الدراسي، وذلك بغية المزيد من الوضوح فيما نقدمه.

وقد نُشرت الفصول الثلاثة عشر الأولى من الطبعة الثالثة لـ" **نماذج إحصائية خطّية تطبيقية**" في كتاب منفصل تحت عنوان" **نماذج انحدار خطية تطبيقية**"، طبعة ثانية. ويتضمن الكتاب الأخير هذا ثلاثة فصول إضافية هي تحليل الارتباط (الفصل ١٤)، الانحدار غير الخطي (الفصل ١٥) وتقنيات الانحدار عندما يكون المتغير المستقل ثنائيا (الفصل ١٦).

وإحدى الميزات الرئيسة للطبعة الثالثة من نماذج إحصائية خطية تطبيقية هو الأسلوب الموحد لتطبيق نماذج إحصائية خطية في الانحدار، وفي تحليل التباين، وفي التصاميم التجريبية. وبدلا من معالجة هذه الميادين بصورة منعزلة فإننا نسعى إلى تبيان العلاقات الضمنية بينها واستخدام رموز مشتركة في الانحدار، من جهة، وفي تحليل التباين والتصاميم التجريبية من جهة أخرى، يسهّل النظرة الموحدة لها جميعا. وقد نُقلت فكرة النموذج الإحصائي الخطي العام، والتي تبرز بصورة طبيعية في سياق نماذج الانحدار، إلى نماذج تحليل التباين ونماذج التصاميم التجريبية، كي تُظهر علاقتها بنماذج الانحدار. ولهذا الأسلوب الموحد أيضا ميزة البساطة في العرض.

و لم يشتمل هذا الكتاب فقط على المواضيع الأكثر تقليدية في الانحدار وتحليل التباين والتصاميم التجريبية الأساسية، ولكنه تطرق أيضا لمواضيع، كثيرا ماأستُخفَّت مع أنها مهمة في الممارسةالعملية. وهكذا فقد كرسنا فصلا بكامله (الفصل العاشر) لمتغيرات مؤشرة مستقلة. وينبري فصل آخر (الفصل ١٢) إلى عملية بناء نموذج انحدار، بما في ذلك طرق اختيار بمساعدة الحاسوب لتحديد مجموعات جزئية "جيدة" من المتغيرات المستقلة وتحليلها تحليلا شاملا قبل القيام بالاختيار النهائي لنموذج الانحدار، ومن ثمَّ التحقق من صحة نموذج الانحدار المختار. واستخدام تحليل الراسب وتشخيصات أخرى لفحص مصداقية نموذج انحدار هو إيقاع متواتر عبر هذا الكتاب. وكذلك الأمر بالنسبة لاستخدام تدابير علاجية يمكن أن تكون مفيدة عندما لايكون النموذج مناسبا. ونؤكد، في تحليل نتائج دراسة، على استخدام طرق التقدير أكثر من اختبارات المعنوية، لأن التقدير، في الغالب، أكثر مغزى في الممارسة العملية وبما أنه من النادر أن تُعنى المسائل التطبيقية بتقدير بمفرده فقد أكدنا أيضا على استخدام طرق التقدير المتزامن.

وقد قُدّمت الأفكار النظرية إلى الدرجة التي نحتاجها من أجل فهم رشيد عند القيام بتطبيقات سليمة. وأعطيت البراهين في ظروف نشعر معها أنها تخدم في إيضاح طريقة عمل. وجرى التأكيد على فهم شامل للنماذج، وعلى وجه الخصوص فهم معنى معالم النموذج. ذلك لأن مثل هذا الفهم أمر أساسي لسلامة التطبيقات. ويتضمن الكتاب تشكيلة واسعة من الأمثلة الواقعية وذلك لتوضيح استخدام المبادىء النظرية، ولتبيان التنوع العظيم لتطبيقات النماذج الإحصائية الخطية، ولإظهار كيفية القيام \التحاليل في المسائل المختلفة.

ونستخدم فقرات تحت عنوان " ملاحظات" أو " تعليقات " في كل صل لتقديم مناقشة إضافية ومسائل تتصل بالمجرى الرئيس لتطور النقاش، وبهذه الطريقة يبقى تقديم الأفكار الأساسية في الفصل تقديما يتلافى التفاصيل والمنعطفات التي قد تصرف القارىء عن الفكرة الأساسية.

وكثيرا ماتتطلب تطبيقات النماذج الإحصائية الخطية حسابات مستفيضة. وننطلق من موقع أن الحاسوب متوافر في معظم العمل التطبيقي، وفضلا عن ذلك ففي متناول كل مستخدم للحاسوب أنواع مختلفة من الحزم البرامجية الخاصة بتحليل الانحدار وتحليل التباين. وبالتالي فإننا نشرح الخطوات الرياضية الأساسية في توفيق نموذج إحصائي خطي دون الإسهاب في التفاصيل الحسابية. ويسمح لنا هذا الأسلوب بتجنب العديد من الصيغ المعقدة، ونستطيع معه التركيز على المبادىء الأساسية. ونستخدم في هذا الكتاب المدرسي قدرات الحاسوب على إنجاز الحسابات استخداما واسعا، ونوضح تشكيلة من مُخرجات الحاسوب شارحين كيفية استخدامها في التحليل.

وفي نهاية كل فصل (باستثناء الفصل الأول) نقدم مختارات من المسائل. ويمكن للقارىء هنا أن يعزز فهمه للطرائقية ويستخدم المفاهيم التي تعلمها في تحليل البيانات. وقد حرصنا على تقديم مسائل تحليل بيانات تمثل تطبيقات أصيلة. وأفضل طريقة للقيام بالحسابات في معظم المسائل هي استخدام حاسب يدوي أو حاسب آلي (حاسوب).

ونفترض أن قارىء الطبعة الثالثة من نماذج إحصائية خطية تطبيقية قد اجتاز مقررا، يشكّل مدخلا إلى الاستقراء الاحصائي، ويغطي المادة التي أوجزناها في الفصل الأول.

وحساب التفاضل والتكامل غير مطلوب لقراءة نماذج إحصائية خطية تطبيقية ونستخدم أحيانا حساب التفاضل والتكامل لتبيان كيفية الحصول على بعض النتائج المهمة، إلا أن هذه الإثباتات مقصورة على التعليقات أو الملاحظات الإضافية ويمكن حذفها دون أية خسارة في استمرارية دراسة الكتاب. وسيجد القرّاء ذوو المعرفة بحساب التفاضل والتكامل هذه التعليقات والملاحظات في تسلسلها الطبيعي بحيث يحصلون على فوائد المعالجات الرياضية في سياقها المباشر وفي النماذج الخطية بصورة عامة، وفي الانحدار المتعدد على وجه الخصوص، نحتاج الى بعض العناصر الأساسية من جبر المصفوفات ويقدم الفصل السادس هذه العناصر من جبر المصفوفات في سياق الانحدار البسيط تسهيلا لتعلمها.

والطبعة الثالثة من نماذج احصائية خطية تطبيقية مصممة لاستخدامها في مقررات في النماذج الإحصائية الخطية من مستوى المرحلة الجامعية الأولى ومن مستوى الدراسات العليا، وكمقررات ثانية في الاحصاء التطبيقي. ويعتمد مدى استخدام المادة المقدمة في هذا الكتاب المدرسي في مقرر معين على مقدار الوقت المتوفر وعلى اهداف المقرر. وبعض من المقررات الممكنة تشمل:

١- مقرر لفصلين دراسيين، كل منهما نصف سنوي، أو لفصلين دراسيين كل منهما ثلث سنوي، في الانحدار، تحليل التباين والتصاميم التجريبية الأساسية يمكن أن يبنى على الفصول التالية:

الانحدار: ٢ ، ٣ ، ٤ ، ٥ (الفقرات من (٥-١)إلى (٥-٣))، ٦ ، ٧ ، ٨ ، ١٠ (الفقرات من (١٠-١) إلى(١٠-٤)) ، ١١(الفقرات من (١١-١) إلى (١١-٦))، ١٢.

تحليل التباين : ١٤ ، ١٥ ، ١٦ ، ١٨ ، ١٩ ، ٢٠ .

تصاميم تجريبية : ٢٤ ، ٢٥ ، ٢٦ ، ٢٩ .

٢- يمكن أن يبنى مقرر، لفصل ثلثي (Quarter) أو لفصل نصفي (Term)، في تحليل الانحدار على الفصول التالية ٢، ٣، ٤، ٥ (الفقرات من (٥-١) إلى(٥-٣))، ٦، ٧ ، ٨ ، ٩، ١٠ (الفقرات من (١٠-١) إلى (١٠-٤))، ١١(مواضيع مختارة)، ١٢، ١٣.

٣- يمكن أن يبنى مقرر، لفصل ثلثي أو لفصل نصفي، في تحليل التباين على الفصول التالية: ١٤،١٥،١٦، ١٧ (مواضيع مختارة)، ١٨، ١٩، ٢٠ ، ٢١ (مواضيع مختارة)،٢٢، ٢٣.

٤- يمكن أن يُبنى مقـرر، لفصـل ثلثي أو لفصـل نصفـي، في الانحـدار وتحليـل التبـاين علـى الفصول التالية:

الانحدار: ٢،٣،٤،٥ (الفقرات من(٥-١) إلى (٥-٣))، ٦، ٧،٨ ، ١٠ (الفقرات من (١٠-١) إلى (١٠-٤)).

تحليل التباين: ١٤ ، ١٥ ، ١٦ ، ١٨، ١٩.

٥- يمكن أن يبنى مقرر، لفصل ثلثي أو لفصل نصفي، في التصاميم التجريبية الأساسية علـى الفصول التالية: ٢٤، ٢٥، ٢٦، ٢٧، ٢٨، ٢٩.

وبالقدر الذي يسمح به الوقت يمكن للمدرس أن يغطي مواضيع إضافية من الكتاب.

ويمكن استخدام هذا الكتاب أيضا في دراسة شخصية لأشخاص يهتمـون بميـادين إدارة الأعمال، الاقتصاد، والعلوم الاجتماعيـة، الصحيـة والأحيائيـة، ممـن يرغبـون في تحصيـل كفاءة في تطبيق النماذج الإحصائية الخطية.

ويمكن للمدرسين الحصول على كتيّب الحلول من الناشر، إروين (Irwin). ويتضمن هذا الكتيّب، على قرص (ديسكت)، البيانـات لجميـع المسـائل والتمـارين والمشـاريع، ومجموعـات البيانات في الملحق.

ولايمكن تأليف كتاب كهذا دون مساعدة كبيرة من آخرين. ونحن مدينون للعديـد ممـن ساهموا في تطويـر النظريـة والتطبيقـات الـتي نوقشـت في هـذا الكتـاب. ونحـب أيضـا التنويـه بإعجابنا بطلابنا الذين ساعدونا بمختلف الطرق على تحديث طريقة العـرض في هـذا الكتـاب. وممنونون للعديد من مستخدمي نماذج إحصائية خطية تطبيقيـة ونمـاذج انحـدار خطيـة تطبيقيـة الذين زودونا بتعليقاتهم ومقترحاتهم النابعة من تدريسهم لهذين الكتابين. ونحن مدينون أيضـا للأساتذة جيمس هولسـتاين (James E. Holstein) جامعـة ميسـوري (Missouri)، وديفيـد شيري(David L . Sherry) جامعة غرب فلوريدا (West Florida)، لمراجعتهم الطبعـة الأولى لنماذج إحصائية خطية تطبيقية، وللأساتذة صموئيل كوتز (Samuel Kotz) جامعـة مـيريلاند (Mary Land)، رالـف روسـو(Ralph P. Russo) جامعـة أيـووا(Iowa)، وبيـــتر ثـــال

(Peter F. Thall) جامعة جورج واشنطن (George Washington) لمراجعتهم كتاب نماذج انحدار خطية تطبيقية، وللأساتذة جون شيو(John S. Y. Chiu) جامعة واشنطن، وجيمس كالفين (James A. Calvin) جامعة ايووا، وميخائيل دريسكول (Michael F. Oriscoll) جامعة ولاية أريزونا (Arizona State) لمراجعتهم الطبعة الثانية من نماذج احصائية خطية تطبيقية. ولقد قدم هؤلاء المراجعون العديد من المقترحات المهمة، التي تستحق جزيل امتناننا.

وقد ساعدنا جورج كاتسونيس(George Cotsonis)، مارجريت كولشاك (Margarette S. Kolczak) وآلفين رامبي(Alvin H. Rampey) بشكل متقن في تدقيق المخطوطة، وفي إعداد الرسوم باستخدام الحاسوب، وبطرق أخرى. أما جين ديزني (June Disney) وساندرا جون هاتفيلد (Sandra June Hatfield) فقد قامتا بجميع الجهد الطباعي تقريبا، وتصدتا بمقدرة لتهيئة مخطوطة صعبة. ونحن ممتنون جدا لهؤلاء الأشخاص جميعا لعونهم ومساعدتهم.

وأخيرا فقد تحملت عائلاتنا بصبر، الضغوط التي سببها التزامنا باستكمال هذه النسخة المنقحة، ونحن مقدّرون لتسامحهم.

John Neter	جون نتر
William Wasserman	ويليام وازرمان
Michael H. Kutner	ميخائل كتنر



المحتويات

الفصل الأول: بعض النتائج الأساسية في الاحتمال والإحصاء

الباب الأول: الانحدار الخطي البسيط

الفصل الثاني: الانحدار الخطي بمتغير مستقل واحد

الفصل الثالث: استقراءات في تحليل الانحدار

الفصل الرابع: تشخيصات وتدابير علاجية

الفصل الخامس: استقراءات متزامنة ومواضيع أخرى في تحليل الانحدار

الفصل السادس: طريقة المصفوفة لتحليل الانحدار الخطي البسيط

الباب الثاني: انحدار خطي عام

الفصل السابع: الانحدار المتعدد - I

الفصل الثامن: الانحدار المتعدد- II

الفصل التاسع: انحدار كثيرات الحدود

الفصل العاشر: المتغيرات المستقلة النوعية

الفصل الحادي عشر: تشخيصات وتدابير علاجية - II

الفصل الثاني عشر: بناء نموذج الانحدار

الفصل الثالث عشر: الارتباط الذاتي في بيانات السلاسل الزمنية

الملاحق

ثبت المصطلحات

بعض النتائج الأساسية في الإحصاء والاحتمال

يتضمن هذا الفصل بعض النتائج الأساسية في الإحصاء والاحتمال. وهو معد كفصل مرجعي يمكنك الرجوع إليه عند قراءتك للكتاب. وفي سياق الكتاب نشير أحيانا إشارات محددة إلى نتائج من هذا الفصل. وفي أحيان أخرى قد ترغب العودة بنفسك إلى نتائج معينة من هذا الفصل، وذلك عندما تشعر بالحاجة إلى ذلك.

وقد تفضل استعراض نتائج الاحتمال والاستقراء الإحصائي في هذا الفصل قبل قراءة الفصل الثاني أو يمكنك المضي مباشرة إلى الفصل التالي.

(١-١) مؤثرا الجمع والضرب

مؤثر الجمع

يُعرَّف مؤثر الجمع Σ كما يلي:

$$\sum_{i=1}^{n} Y_i = Y_1 + Y_2 + \ldots + Y_n \qquad (1.1)$$

بعض الخصائص المهمة لهذا المؤثر هي:

$$\sum_{i=1}^{n} k = nk \qquad (1.2a)$$

حيث k ثابت.

$$\sum_{i=1}^{n}(Y_i + Z_i) = \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} Z_i \qquad (1.2b)$$

$$\sum_{i=1}^{n}\left(a + cY_i\right) = na + c\sum_{i=1}^{m} Y_i \qquad (1.2c)$$

حيث a و c ثابتان.

ويُعرف مؤثر الجمع المضاعف $\Sigma\Sigma$ كما يلي:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij} = \sum_{i=1}^{n}\left(Y_{i1} + \ldots + Y_{im}\right) \qquad (1.3)$$
$$= Y_{11} + \ldots + Y_{1m} + Y_{21} + \ldots + Y_{2m} + \ldots + Y_{nm}$$

وإحدى الخواص المهمة لمؤثر الجمع المضاعف.

$$\sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} = \sum_{j=1}^{m} \sum_{i=1}^{n} Y_{ij} \qquad (1.4)$$

مؤثر الضرب

يعرّف مؤثر الضرب $\prod$ كما يلي:

$$\prod_{i=1}^{n} Y_i = Y_1 . Y_2 . Y_3 \cdots Y_n \qquad (1.5)$$

(١-٢) الاحتمال

نظرية الجمع

لتكن A_i و A_j حادثتين معرفتين في فضاء عينة، فعندئذ:

$$P(A_i \cup A_j) = P(A_i) + P(A_j) - P(A_i \cap A_j) \qquad (1.6)$$

حيث يرمز $P(A_i \cup A_j)$ لاحتمال وقوع A_i أو A_j أو وقوعهما معا. ويرمز $P(A_i)$ و $P(A_j)$ لاحتمال A_i واحتمال A_j، على الترتيب، ويرمز $P(A_i \cap A_j)$ لاحتمال وقوع كل من A_i و A_j معا.

نظرية الضرب

لنرمز بـ $P(A_i \mid A_j)$ للاحتمال الشرطي لوقوع A_i علما أن A_j قد وقعت. يُعرّف هذا الاحتمال الشرطي كما يلي:

$$P(A_i \mid A_j) = \frac{P(A_i \cap A_i)}{P(A_i)} \qquad P(A_i) \neq 0 \qquad (1.7)$$

وتعرضُ نظرية الضرب أن:

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j \mid A_i) \qquad (1.8) \\ &= P(A_j)P(A_i \mid A_j) \end{aligned}$$

الحوادث المتممة

نستخدم الرمز $\overline{A_i}$ للإشارة لمتمم الحادثة A_i. وتُعتبر النتائج التالية حول الحوادث المتممة مفيدة:

$$P(\overline{A_i}) = 1 - P(A_i) \qquad (1.9)$$

$$P(\overline{A_i \cup A_j}) = P(\overline{A_i} \cap \overline{A_j}) \qquad (1.10)$$

(١-٣) المتغيرات العشوائية

نفترض، عبر هذه الفقرة، أن المتغير العشوائي Y يأخذ عددا منتهيا من القيم، ما لم يُذكر خلاف ذلك.

القيمة المتوقعة

لنفترض أن المتغير العشوائي Y يأخذ القيم Y_1، Y_2،...، Y_k باحتمالات معطاة بدالة الاحتمال:

$$f(Y_s) = P(Y = Y_s) \qquad s = 1,\ldots, k \tag{1.11}$$

تعرّف قيمة توقع Y ويُرمز لها بـ $E\{Y\}$ كما يلي:

$$E\{Y\} = \sum_{S=1}^{k} Y_s f(Y_s) \tag{1.12}$$

ويُسمّى $E\{\ \}$ مؤثر التوقع.

وإحدى الخواص المهمة لمؤثر التوقع E هي:

$$E\{a + cY\} = a + cE\{Y\} \tag{1.13}$$

حيث a وc ثابتان. وكحالات خاصة من هذه نجد:

$$E\{a\} = a \tag{1.13a}$$

$$E\{cY\} = cE\{Y\} \tag{1.13b}$$

$$E\{a + Y\} = a + E\{Y\} \tag{1.13c}$$

ملاحظة

إذا كان Y متغيرا عشوائيا متصلا دالة كثافته $f(Y)$، فيعرّف $E\{Y\}$ كما يلي:

$$E\{Y\} = \int_{-\infty}^{\infty} Yf(Y)\, dY \tag{1.14}$$

التباين

يعرّف تباين المتغير العشوائي Y، ويرمز له بالرمز $\sigma^2\{Y\}$ كما يلي:

$$\sigma^2\{Y\} = E\left\{\left(Y - E\{Y\}\right)^2\right\} \tag{1.15}$$

والعبارة التالية هي عبارة مكافئة:

$$\sigma^2\{Y\} = E\left\{Y^2\right\} - \left(E\{Y\}\right)^2 \tag{1.15a}$$

ويسمى $\sigma^2\{\ \}$ مؤثر التباين.

كثيرا ما نحتاج إلى حساب تباين دالة خطية في Y وباستخدام $\sigma^2\{a+cY\}$ كرمز لتباين $a + cY$ نجد:

$$\sigma^2\{a + cY\} = c^2\sigma^2\{Y\} \tag{1.16}$$

حيث a و c ثابتان.

وكحالات خاصة من هذه النتيجة نجد:

$$\sigma^2\{a + Y\} = \sigma^2\{Y\} \tag{1.16a}$$

$$\sigma^2\{cY\} = c^2\sigma^2\{Y\} \tag{1.16b}$$

ملاحظة

إذا كان Y متصلا فإن $\sigma\{Y\}$ يُعرّف كما يلي:

$$\sigma^2\{Y\} = \int_{-\infty}^{\infty}\left(Y - E\{Y\}\right)^2 f(Y)\,dY \tag{1.17}$$

التوزيعات الاحتمالية المشتركة، الهامشية والشرطية

لنرمز بـ $g(Y,z)$ لدالة الاحتمال المشتركة للمتغيرين العشوائيين Y و z:

$$g(Y_s, Z_t) = P(Y = Y_s \cap Z = Z_t) \qquad s = 1,\ldots,k;\ t = 1,\ldots,m \tag{1.18}$$

فدالة الاحتمال الهامشية للمتغير Y، ونرمز لها بـ $f(Y)$، هي:

$$f(Y_s) = \sum_{t=1}^{m} g(Y_s, Z_t) \qquad s = 1,\ldots, k \tag{1.19a}$$

ودالة الاحتمال الهامشية للمتغير z، ونرمز لها بـ $h(Z)$، هي:

$$h(Z_t) = \sum_{s=1}^{k} g(Y_s, Z_t) \qquad ;\ t = 1,\ldots, m \tag{1.19b}$$

أما دالة الاحتمال الشرطية للمتغير Y علما أن $Z = Z_t$ فهي:

$$f(Y_s|Z_t) = \frac{g(Y_s, Z_t)}{h(Z_t)} \qquad ;\ h(Z_t) \neq 0;\ s = 1,\ldots,k \tag{1.20a}$$

وتكون دالة الاحتمال الشرطية للمتغير Z علما أن $Y = Y_s$:

$$h(Z_t|Y_s) = \frac{g(Y_s, Z_t)}{f(Y_s)} \qquad f(Y_s) \neq 0;\ t = 1,\ldots,m \tag{1.20b}$$

التغاير

يعرّف تغاير Y و Z، ويرمز له بـ $\sigma\{Y,Z\}$، كما يلي:

$$\sigma\{Y, Z\} = E\left\{\left(Y - E\{Y\}\right)\left(Z - E\{Z\}\right)\right\} \tag{1.21}$$

وبتعبير مكافيء:

$$\sigma\{Y,Z\} = E\{YZ\} - \big(E\{Y\}\big)\big(E\{Z\}\big) \tag{1.21a}$$

يرمز $\sigma(a_1 + c_1Y, a_2 + c_2 Z)$ لتغاير $a_1 + c_1Y$ و $a_2 + c_2Z$ ولدينا:

$$\sigma\{a_1 + c_1\, Y, a_2 + c_2Z\} = c_1c_2\sigma\{Y, Z\} \tag{1.22}$$

حيث a_1، a_2، c_1 و c_2 ثوابت.

وكحالات خاصة من هذه نجد:

$$\sigma\{c_1\, Y, c_2Z\} = c_1c_2\sigma\{Y, Z\} \tag{1.22a}$$

$$\sigma\{a_1 + Y, a_2 + Z\} = \sigma\{Y, Z\} \tag{1.22b}$$

ولدينا من التعريف:

$$\sigma\{Y, Y\} = \sigma^2\{Y\} \tag{1.23}$$

حيث $\sigma^2\{Y\}$ تباين Y.

المتغيرات العشوائية المستقلة

يكون المتغيران العشوائيان Y و Z مستقلين إذا، وفقط إذا، كان:

$$g\{Y_s\,, Z_t\} = f(Y_s)\, h(Z_t) \qquad s = 1,\ldots,k;\ t = 1,\ldots,m \tag{1.24}$$

وإذا كان Y و Z متغيرين عشوائيين مستقلين فإن:

$$\sigma\{Y, Z\} = 0 \qquad \text{عند استقلال } Y \text{ و } Z \tag{1.25}$$

(في الحالة الخاصة عندما يتوزع Y و Z وفق توزيع مشترك طبيعي فإن: $\sigma\{Y,\ Z\} = 0$ يؤدي إلى استقلال Y و Z).

دوال في متغيرات عشوائية

ليكن $Y_1,\ldots,Y_n$ عدد n من المتغيرات العشوائية ولنعتبر الدالة $\sum a_i\, Y_i$ حيث a_i ثوابت. فلدينا عندئذ:

$$E\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n} a_i\; E\{Y_i\} \tag{1.26a}$$

$$\sigma^2\left\{\sum_{i=1}^{n} a_i Y_i\right\} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma\{Y_i, Y_j\} \tag{1.26b}$$

حيث a_i ثوابت.

وبالتحديد لدينا في حالة $n = 2$:

$$E\{a_1Y_1 + a_2Y_2\} = a_1E\{Y_1\} + a_2\{Y_2\} \tag{1.27a}$$

$$\sigma^2\{a_1Y_1 + a_2Y_2\} = a_1^2\sigma^2\{Y_1\} + a_2^2\sigma^2\{Y_2\} + 2a_1a_2\sigma\{Y_1, Y_2\} \tag{1.27b}$$

وإذا كانت المتغيرات العشوائية Y_i مستقلة فلدينا:

$$\sigma^2\left\{\sum_{i=1}^{n} a_iY_i\right\} = \sum_{i=1}^{n} a_i^2\sigma^2\{Y_i\} \tag{1.28}$$ عندما تكون Y_i مستقلة

وكحالات خاصة من هذه نجد:

$$\sigma^2\{Y_1 + Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} \tag{1.28a}$$ عند استقلال Y_1, Y_2

$$\sigma^2\{Y_1 - Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} \tag{1.28b}$$ عند استقلال Y_1, Y_2

وعند استقلال المتغيرات العشوائية Y_i يعطى تغاير دالتين خطيتين $\sum a_i Y_i$ و $\sum c_i Y_i$ بالعلاقة:

$$\sigma\left\{\sum_{i=1}^{n} a_iY_i, \sum_{i=1}^{n} c_iY_i\right\} = \sum_{i=1}^{n} a_ic_i\sigma^2\{Y_i\} \tag{1.29}$$ عند استقلال Y_i

نظرية النهاية المركزية

إذا كانت $Y_1, ..., Y_n$ مشاهدات عشوائية مستقلة من مجتمع دالة احتماله $f(Y)$ وتباينه $\sigma^2\{Y\}$ محدود، فإن توزيع متوسط العينة $\overline{Y}$:

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \tag{1.30}$$

هو على وجه التقريب توزيع طبيعي متوسطه $E\{Y\}$ وتباينه $\sigma^2\{Y\}/n$، وذلك عندما يكون n كبيرا بما فيه الكفاية.

(١-٤) التوزيع الاحتمالي الطبيعي والتوزيعات المتعلقة به.

التوزيع الاحتمالي الطبيعي

تعطى دالة كثافة متغير عشوائي طبيعي بالعلاقة:

$$f(Y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{Y-\mu}{\sigma}\right)^2\right] \quad ; -\infty < Y < +\infty \tag{1.31}$$

حيث μ و σ معلمتا التوزيع و $exp\ (a)$ تعني e^a.

متوسط وتباين متغير عشوائي طبيعي Y هما:

$$E\{Y\} = \mu \quad (1.32a)$$
$$\sigma^2\{Y\} = \sigma^2 \quad (1.32b)$$

دالة في متغير عشوائي طبيعي. تمتلك الدالة الخطية في متغير عشوائي طبيعي الخاصة التالية:

(1.33) إذا كان Y متغيرا عشوائيا طبيعيا، فإن المتغير المحوّل $Y' = a + cY$

(حيث a و c ثابتان) يتوزع طبيعيا بمتوسط $a + cE\{Y\}$ وتباين $c^2\sigma^2\{Y\}$.

متغير طبيعي معياري. يتبع المتغير الطبيعي المعياري z:

(1.34) $z = \dfrac{Y-\mu}{\sigma}$ حيث Y متغير عشوائي طبيعي له

توزيع طبيعي متوسطه 0 وتباينه 1. ونشير إلى هذا كما يلي:

(1.35) z يتبع $N\ (0,1)$

متوسط ← 0 ، تباين ← 1

يتضمن الجدول (أ-١) في الملحق أ الاحتمالات التراكمية A للمئينات $z(A)$ حيث:

$$P\{z \le z(A)\} = A \quad (1.36)$$

فمثلا، عندما $z(A) = 2.00$ يكون $A = 0.9772$. ولأن التوزيع الطبيعي المعياري متناظر حول الصفر فإن $z(A) = -2.00$ تقابل $A = 1\text{-}0.9772 = 0.0228$.

دالة في متغيرات عشوائية طبيعية مستقلة. لتكن $Y_1, ..., Y_n$ متغيرات عشوائية طبيعية ومستقلة، فلدينا عندئذ القاعدة التالية:

(1.37) عندما تكون $Y_1, ..., Y_n$ متغيرات عشوائية طبيعية مستقلة يكون توزيع التركيب الخطي $a_1Y_1 + a_2Y_2 + ... + a_nY_n$ توزيعا طبيعيا بمتوسط $\sum a_i E\{Y_i\}$ وتباين $\sum a_i^2 \sigma^2\{Y_i\}$.

توزيع χ^2

لنفرض $z_1, ..., z_\nu$ من المتغيرات الطبيعية المعيارية والمستقلة، فنعرّف عندئذ:

$$\chi^2(\nu) = z_1^2 + z_2^2 + ... + z_\nu^2 \quad (1.38)$$

حيث المتغيرات z_i مستقلة.

ولتوزيـع χ^2 معلمـة واحـدة تسـمى درجـات الحريـة (*d.f*). ومتوسـط توزيـع χ^2 بدرجات حرية ν هو:

$$E\left\{\chi^2(\nu)\right\} = \nu \tag{1.39}$$

ويحوي الجدول (أ ـ ٣) في الملحـق أ مئينـات عديـد مـن توزيعـات χ^2. ونعـرّف $\chi^2(A;\nu)$ كما يلي:

$$P\left\{\chi^2(\nu) \le \chi^2(A;\nu)\right\} = A \tag{1.40}$$

وبفرض $\nu = 5$ يكون المئين 90 لتوزيع χ^2 بـ 5 درجات حرية $\chi^2(.90;5) = 9.24$

توزيع t

ليكن z و $\chi^2(\nu)$ متغيرين عشوائين مستقلين فنعرف عندئذ:

$$t(\nu) = \frac{z}{\left[\frac{\chi^2(\nu)}{\nu}\right]^{1/2}} \tag{1.41}$$

حيث z و $\chi^2(\nu)$ مستقلان.

وللتوزيع t معلمة واحدة هي درجات الحرية ν. ومتوسـط توزيـع t بـ ν مـن درجات الحرية هو:

$$E\{t(\nu)\} = 0 \tag{1.42}$$

ويحوي الجدول (أ ـ٣) في الملحق أ مئينات عديد من توزيعات t. ونعرّف $t(A;\nu)$ كما يلي:

$$P\{t(\nu) \le t(A;\nu)\} = A \tag{1.43}$$

وبفرض $\nu = 10$ يكون المئين 90 لتوزيع t بـ 10 درجات حرية هـو $t(.90;10)=1.372$.
وحيث إن التوزيع t متناظر حول الصفر فإن $t(.10;10) = -1.372$.

توزيع F

ليكن $\chi^2(\nu_1)$ و $\chi^2(\nu_2)$ متغيري χ^2 مستقلين فنعرّف عندئذ:

$$F(\nu_1,\nu_2)=\frac{\chi^2(\nu_1)}{\nu_1}\div\frac{\chi^2(\nu_2)}{\nu_2} \quad (1.44)$$

حيث $\chi^2(\nu_1)$ و $\chi^2(\nu_2)$ مستقلان.

البسط df — المقام df

وللتوزيع F معلمتان هما درجات حرية البسط ودرجات حرية المقام، وهما هنا ν_1 و ν_2 على الترتيب ويحوي الجدول (أ-٤) في الملحق أ مئينات عديد من توزيعات F.

ونعرف $F(A;\nu_1,\nu_2)$ كما يلي:

$$P\{F(\nu_1,\nu_2)\leq F(A;\nu_1,\nu_2)\}=A \quad (1.45)$$

وبفرض أن $\nu_1=2$ و $\nu_2=3$، يكون المئين 90 لتوزيع بدرجات حرية 2 و 3 للبسط والمقام، على الترتيب $F(0.90;2,3)=5.46$.

يمكن الحصول على مئينات تحت 50 بالمائة بالاستفادة من العلاقة:

$$F(A;\nu_1,\nu_2)=\frac{1}{F(1-A;\nu_2,\nu_1)} \quad (1.46)$$

وهكذا يكون: $F(.10;3,2)=1/F(.90;2,3)=1/5.46=0.183$

يرتبط المتغيران العشوائيان t و F بالعلاقة التالية:

$$[t(\nu)]^2=F(1,\nu) \quad (1.47a)$$

كما ترتبط مئينات التوزيعين t و F بالعلاقة التالية:

$$[t(.5+A/2;\nu)]^2=F(A;1,\nu) \quad (1.47b)$$

ملاحظة

عبر هذا الكتاب نعتبر $z(A)$، $\chi^2(A;\nu)$، $t(A;\nu)$ و $F(A;\nu_1,\nu_2)$ كمئينات $A(100)$. أو يمكن اعتبارها بصورة مكافئة كَسْريات A.

(١-٥) التقدير الإحصائي

خواص المقدِّرات

(1.48) يقال عن مقدِّر $\hat{\theta}$ للمعلمة θ إنه غير منحاز إذا كان:

$$E\{\hat{\theta}\}=\theta$$

(1.49) و $\hat{\theta}$ مقدِّر متّسق للمعلمة إذا كان:

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right|\geq\varepsilon\right)=0 \quad \text{لكل} \quad \varepsilon>0$$

(1.50) و $\hat{\theta}$ مقدِّر كاف للمعلمة θ إذا كانت دالة الاحتمال المشتركة الشرطية لمشاهدات العينة (علما أن $\hat{\theta}$ مثبتة) لا تعتمد على المعلمة θ.

(1.51) و $\hat{\theta}$ مقدِّر أصغري التباين للمعلم θ إذا كان من أجل أي تقدير آخر $\hat{\theta}*$.

$$\sigma^2\{\hat{\theta}\}\leq\sigma^2\{\hat{\theta}*\} \quad \text{لكل قيم} \quad \hat{\theta}*$$

تقديرات الإمكانية العظمى

طريقة الإمكانية العظمى هي طريقة عامة لإيجاد مقدرات، فلنفرض أننا نعاين مجتمعا دالة احتماله $f(Y;\theta)$ تحوي معلمة واحدة، θ. إذا كانت المشاهدات $Y_1,...,Y_n$ مستقلة فإن دالة الاحتمال المشتركة لمشاهدات العينة هي:

$$g(Y_1,...,Y_n)=\prod_{i=1}^{n} f(Y_i;\theta) \tag{1.52a}$$

وعندما ننظر إلى دالة الاحتمال المشتركة هذه كدالة في θ، مع قيم معطاة للمشاهدات، فإن هذه الدالة تسمى دالة الإمكانية $L(\theta)$.

$$L(\theta)=\prod_{i=1}^{n} f(Y_i;\theta) \tag{1.52b}$$

ومقدِّر الإمكانية العظمى لِ θ هو قيمة θ التي تجعل $L(\theta)$ أعظم مايمكن، وتحت شروط عامة تماما تكون مقدرات الإمكانية العظمى متسقة وكافية.

مقدِّرات المربعات الدنيا

طريقة المربعات الدنيا هي طريقة عامة أخرى لإيجاد مقدِّرات. وفيها نفترض أن مشاهدات العينة تأخذ الشكل (في حالة معلمة واحدة θ)

$$Y_i=f_i(\theta)+\varepsilon_i \qquad i=1,...,n \tag{1.53}$$

حيث $f_i(\theta)$ دالة معروفة في المعلمة θ و ε_i متغيرات عشوائية نفترض، عادة، أن توقعها $E\{\varepsilon_i\}=0$.

وفي طريقة المربعات الدنيا، ومن أجل مشاهدات معطاة للعينة، نعتبر مجموع المربعات:

$$Q=\sum_{i=1}^{n}\left[Y_i-f_i(\theta)\right]^2 \quad (1.54)$$

دالة في θ. ونحصل على مقدر المربعات الدنيا لـ θ بحساب قيمة θ التي تجعل Q أصغر مايمكن. وفي كثير من الحالات تكون مقدرات المربعات الدنيا غير منحازة ومتسقة.

(١-٦) استقراءات حول متوسط مجتمع – مجتمع طبيعي

لدينا عينة عشوائية $Y_1 ,..., Y_n$ حجمها n من مجتمع طبيعي متوسطه μ وانحرافه المعياري σ. فمتوسط العينة وانحرافها المعياري هما:

$$\overline{Y}=\frac{\sum_i Y_i}{n} \quad (1.55a)$$

$$s=\left[\frac{\sum_i (Y_i-\overline{Y})^2}{n-1}\right]^{1/2}=\left[\frac{\sum_i Y_i^2-\frac{\left(\sum_i Y_i\right)^2}{n}}{n-1}\right]^{1/2} \quad (1.55b)$$

والانحراف المعياري المقدَّر لتوزيع المعاينة لِ $\overline{Y}$ هو:

$$s\{\overline{Y}\}=\frac{s}{\sqrt{n}} \quad (1.55c)$$

وبالتالي فإن:

(1.56) $\dfrac{(\overline{Y}-\mu)}{s\ \{\overline{Y}\}}$ يتوزع وفق t بـ $(n - 1)$ درجة حرية، عندما تكون العينة العشوائية من مجتمع طبيعي.

تقدير الفترة

تستخدم (1.56) للحصول على حدّي ثقة للمعلمة μ بمعامل ثقة $1-\alpha$ وذلك كما يلي:

$$\overline{Y}\pm t(1-\alpha/2;n-1)s\{\overline{Y}\} \quad (1.57)$$

مثال (١). أوجد 95 بالمائة فترة ثقة للمعلمة μ في حالة:

$$n=10 \qquad \overline{Y}=20 \qquad s=4$$

لدينا: $t(.975;9) = 2.262 \qquad s\{\overline{Y}\} = \frac{4}{\sqrt{10}} = 1.265$

. لذا فإن حدّي الثقة هما (1.265) 2.262 ± 20 وبالتالي تكون الـ 95% فترة ثقة لـ μ:

$$17.1 \le \mu \le 22.9$$

اختبارات

باستخدام العلاقة (1.56) تعتمد الاختبارات وحيدة الجانب وثنائية الجانب، المتعلقة بمتوسط المجتمع μ على إحصاءة الاختبار:

$$t^* = \frac{\overline{Y} - \mu_0}{s\{\overline{Y}\}} \tag{1.58}$$

ويحوي الجدول (١-١) قواعد القرار لكل من الحالات الثلاث الممكنة مع إبقاء مخاطرة التورط بخطأ من النوع الأول عند المستوى α .

جدول (١-١) قواعد القرارات لاختبارات حول متوسط مجتمع طبيعي μ.

البدائل	قاعدة القرار
	(أ)
$H_0 : \mu = \mu_0$	إذا كان $\lvert t^* \rvert \le t(1 - \alpha/2; n - 1)$ استنتج H_0
$H_a : \mu \ne \mu_0$	إذا كان $\lvert t^* \rvert > t(1 - \alpha/2; n - 1)$ استنتج H_a
	حيث: $t^* = \frac{\overline{Y} - \mu_0}{s\{\overline{Y}\}}$
	(ب)
$H_0 : \mu \ge \mu_0$	إذا كان $t^* \ge t(\alpha; n - 1)$ استنتج H_0
$H_a : \mu < \mu_0$	إذا كان $t^* < t(\alpha; n - 1)$ استنتج H_a
	(ج)
$H_0 : \mu \le \mu_0$	إذا كان $t^* \le t(1 - \alpha; n - 1)$ استنتج H_0
$H_a : \mu > \mu_0$	إذا كان $t^* > t(1 - \alpha; n - 1)$ استنتج H_a

مثال(٢). اختر بين البدائل:

$$H_0: \ \mu \leq 20$$

$$H_a: \ \mu > 20$$

من أجل α مقيدة عند0.05 و:

$$n = 15 \qquad \overline{Y} = 24 \qquad s = 6$$

لدينا:

$$s\{\overline{Y}\} = \frac{6}{\sqrt{15}} = 1.549$$
$$t(.95;14) = 1.761$$

ومن ثَمَّ فإن قاعدة القرار:

إذا كان $t^* \leq 1.761$ ، استنتج H_0

وإذا كان $t^* > 1.761$، استنتج H_a

وحيث إن $t^* = (24 - 20) / 1.549 = 2.58 > 1.761$ فإننا نستنتج H_a.

مثال(٣). اختر بين البدائل

$$H_0: \mu = 10$$
$$H_a: \mu \neq 10$$

من أجل α مقيدة عند 0.02 و:

$$n = 25 \qquad \overline{Y} = 5.7 \qquad s = 8$$

لدينا:

$$s\{\overline{Y}\} = \frac{8}{\sqrt{25}} = 1.6$$
$$t(.99;24) = 2.492$$

لذا فإن قاعدة القرار:

إذا كان $|t^*| \leq 2.492$، استنتج H_0

إذا كان $|t^*| > 2.492$، استنتج H_a

حيث | | تعني القيمة المطلقة. ولأن:

$|t^*| = |(5.7 - 10) / 1.6| = |-2.69| = 2.69 > 2.492$ فإننا نستنتج H_a.

القيمة -P لنتيجة العينة. تُعرّف القيمة - P لنتيجة عينة بأنها احتمال أن نتيجة العينة ربما كانت أكثر تطرفا من القيمة المشاهدة عندما $\mu = \mu_0$. تعزّز القيم -P الكبيرة H_0 بينما تعزز القيم الصغيرة H_a. ويمكن إجراء اختبار بمقارنة القيمة -P مع قيمة المخاطرة α المحددة ونستنتج H_o إذا كانت القيمة -P أكبر أو تساوي قيمة α المحددة ونستنتج H_a إذا كانت القيمة - P أقل من α.

مثال(٤). في المثال (٢) لدينا $t^* = 2.58$. فتكون القيمة -P لنتيجة هذه العينة الاحتمال $P\{t(14) > 2.58\}$. ونجد من الجدول (أ ــ ٢) أن $t(.985;14)=2.415$ و$t(.990;14) = 2.624$ وبالتالي تكون القيمة - P بين 0.010 و 0.015. في الواقع يمكن تبيان أن هذه القيمة هي 0.011 وهكذا في حالة $\alpha = 0.05$ نقرر لصالح H_a.

مثال(٥). في المثال (٣) لدينا $t^* = -2.69$ ونجد من الجدول (أ - ٢) أن $P\{t(24) < -2.69\}$ تقع بين 0.005 و 0.0075. وفي الواقع يمكن تبيان أن هذه القيمة هي 0.0064. وحيث إن الاختبار ثنائي الجانب وتوزيع t متماثل فإن القيمة -P ثنائية الجانب تكون ضعف القيمة أحادية الجانب أو$2(.0064) = .013$.وهكذا نستنتج في حالة $\alpha = .02$ البديل H_a.

العلاقة بين الاختبارات وفترات الثقة. هناك علاقة مباشرة بين الاختبارات وفترات الثقة. فعلى سبيل المثال يمكن استخدام حدّي فترة الثقة ثنائية الجانب (1.57) لاختبار:

$$H_0: \mu = \mu_0$$
$$H_a: \mu \neq \mu_0$$

إذا وقعت μ_0 ضمن فترة الثقة $(1-\alpha)$ فإن قاعدة القرار ثنائية الجانب في الجدول (١-١)أ، عند مستوى معنوية α، تقود إلى استنتاج H_o والعكس بالعكس. وإذا لم تكن ضمن فترة الثقة فإن قاعدة القرار تقود إلى H_a والعكس بالعكس.

ويوجد تقابل مشابه بين فترات الثقة أحادية الجانب وقواعد القرار ذوات الجانب الواحد.

(١-٧) مقارنات متوسطي مجتمعين – المجتمعات الطبيعية

العينتان مستقلتان

لدينا مجتمعان طبيعيان متوسطاهما μ_1 و μ_2 على الترتيب ولهما الانحراف المعياري σ نفسه. ويراد مقارنة المتوسطين μ_1 و μ_2 بالاستناد إلى عينتين مستقلتين من المجتمعين:

العينة ١: $Y_1, \cdots, Y_{n_1}$

العينة ٢: $Z_1, \cdots, Z_{n_2}$

مقدّرا المتوسطين هما متوسطا العينتين.

$$\overline{Y} = \frac{\sum_i Y_i}{n_1} \tag{1.59a}$$

$$Z = \frac{\sum_i Z_i}{n_2} \tag{1.59b}$$

ويكون $\overline{Y} - \overline{Z}$ مقدّرا لِ $\mu_1 - \mu_2$، ونأخذ:

$$s^2 = \frac{\sum_i (Y_i - \overline{Y})^2 + \sum_i (Z_i - \overline{Z})^2}{n_1 + n_2 - 2} \tag{1.60}$$

مُقدّرا للتباين المشترك $\sigma^2\{\overline{Y} - \overline{Z}\}$ وكمقدّر لتباين الفرق $\overline{Y} - \overline{Z}$ نأخذ:

$$s^2\{\overline{Y} - \overline{Z}\} = s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \tag{1.61}$$

(1.62) وعندما تؤخذ العينتان المستقلتان من مجتمعين طبيعيين لهما الانحراف المعياري نفسه يتوزع $\frac{\left[(\overline{Y} - \overline{Z}) - (\mu_1 - \mu_2)\right]}{s(\overline{Y} - \overline{Z})}$ وفق t بِ $n_1 + n_2 - 2$ درجة حرية.

تقدير فترة. نحصل على حدّي فترة ثقة لِ $\mu_1 - \mu_2$ بمعامل ثقة $1 - \alpha$ عن طريق (1.62):

$$(\overline{Y} - \overline{Z}) \pm t(1 - \alpha/2; n_1 + n_2 - 2)s\{\overline{Y} - \overline{Z}\} \tag{1.63}$$

مثال(٦). أوجد 95 بالمائة فترة ثقة لـ $\mu_1 - \mu_2$ في حالة:

$$n_1 = 10 \quad \overline{Y} = 14 \qquad \Sigma(Y_i - \overline{Y})^2 = 105$$

$$n_1 = 20 \quad \overline{Z} = 8 \qquad \Sigma(Z_i - \overline{Z})^2 = 224$$

لدينا:

$$s^2 = \frac{105+224}{10+20+-2} = 11.75$$

$$s^2\{\overline{Y}-\overline{Z}\} = 11.75\left(\frac{1}{10}+\frac{1}{20}\right) = 1.7625$$

$$S\{\overline{Y}-\overline{Z}\} = 1.328$$

$$t(.975;28) = 2.048$$

$$3.3 = (14 - 8) - 2.048\ (1.328) \le \mu_1 - \mu_2 \le (14 - 8) + 4.048(1.328) = 8.7$$

اختبارات. يمكن وضع اختبارات أحادية الجانب وثنائية الجانب حول μ_1-μ_2 بالاستفادة من (1.62) ويتضمن الجدول (١-٢) قواعد اتخاذ القرار لكل من ثلاث حالات ممكنة. هذه القواعد تستند إلى إحصاء الاختبار:

$$t^* = \frac{\overline{Y}-\overline{Z}}{s\{\overline{Y}-\overline{Z}\}} \quad (1.64)$$

مع إبقاء مخاطرة التورط بخطأ من النوع الأول عند المستوى α .

جدول (١-٢) قواعد القرار لاختبارات حول المتوسطات μ_1 و μ_2 لمجتمعين طبيعيين(σ_1=σ_2=σ)– عيّنات مستقلّة

البدائل	قاعدة القرار
	(أ)
H_0: $\mu_1 = \mu_2$	إذا كان $\|t^*\| \le t(1 - \alpha/2; n_1 + n_2 - 2)$ استنتج H_0
H_a: $\mu_1 \ne \mu_2$	إذا كان $\|t^*\| > t(1 - \alpha/2; n_1 + n_2 - 2)$ استنتج H_a
	حيث: $t^* = \frac{\overline{Y}-\overline{Z}}{s\{\overline{Y}-\overline{Z}\}}$
	(ب)
H_0: $\mu_1 \ge \mu_2$	إذا كان $t^* \ge t(\alpha; n_1 + n_2 - 2)$ استنتج H_0
H_a: $\mu_1 < \mu_2$	إذا كان $t^* < t(\alpha; n_1 + n_2 - 2)$ استنتج H_a
	(ج)
H_0: $\mu_1 \le \mu_2$	إذا كان $t^* \le t(1 - \alpha; n_1 + n_2 - 2)$ استنتج H_0
H_a: $\mu_1 > \mu_2$	إذا كان $t^* > t(1 - \alpha; n_1 + n_2 - 2)$ استنتج H_a

مثال (٧). اختر بين البدائل

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

مقيدا α عند 0.10 ومستخدما بيانات المثال (٦). نحتاج للقيمة $t(.95; 28) = 1.701$، وعليه تكون قاعدة القرار:

إذا كانت $|t^*| \leq 1.701$ قرر H_0

إذا كانت $|t^*| > 1.701$ قرر H_a

وحيث إن $|t^*| = |(14 - 8) / 1.328| = |4.52| = 4.52 > 1.701$، فإننا نقرر لصالح H_a. إن القيمة -P أحادية الجانب هنا هي $P\{t(28) > 4.52\}$. ونرى من الجدول (أ - ٢) أن القيمة -P هذه أقل من 0.0005. في الواقع يمكن تبيان أن هذه القيمة تساوي 0.00005. وبالتالي فإن 0.0001 هي القيمة - P ثنائية الجانب. ومن أجل $\alpha = 0.10$ يكون H_a القرار المناسب.

المشاهدات كأزواج

عندما تكون مشاهدات العينتين عبارة عن أزواج (مثلا Y_i و Z_i قياسان يعكسان موقف المستخدم i من عمله قبل وبعد مضي سنة خبرة، وذلك في عينة من المستخدمين)، نستخدم الفروق:

$$W_i = Y_i - Z_i \quad ; \quad i = 1, \ldots, n \qquad (1.65)$$

كما لو كانت تمثل عينة من مجتمع واحد. وهكذا، عندما نستطيع معاملة W_i كمشاهدات من مجتمع طبيعي لدينا:

(1.66) $\dfrac{[\overline{W} - (\mu_1 - \mu_2)]}{s(\overline{W})}$ يتوزع وفق t بدرجات حرية $n - 1$. وذلك عندما يمكن اعتبار الفروق W_i كمشاهدات من مجتمع طبيعي، حيث:

$$\overline{W} = \frac{\sum_i W_i}{n}$$

$$s^2\{\overline{W}\} = \left(\frac{\sum_i (W_i - \overline{W})^2}{n-1} \right) \div n$$

(١–٨) استقراءات حول تباين مجتمع – المجتمع الطبيعي

في حالة المعاينة من مجتمع طبيعي، تصح العلاقات التالية لتباين العينة s^2: حيث s معرف في (1.55b):

(1.67) $\frac{(n-1)s^2}{\sigma^2}$ يتوزع وفق χ^2 بدرجات حرية $n-1$ عندما تكون العينة العشوائية من مجتمع طبيعي.

تقدير فترة

يمكن الحصول على حد الثقة الأدنى L وحد الثقة الأعلى U لفترة ثقة لتباين المجتمع σ^2 بمعامل ثقة $1 - \alpha$ باستخدام العلاقة (1.67):

$$L = \frac{(n-1)s^2}{\chi^2(1-\alpha/2; n-1)} \qquad U = \frac{(n-1)s^2}{\chi^2(\alpha/2; n-1)} \tag{1.68}$$

مثال (٨). أوجد 98 بالمائة فترة ثقة لِـ σ^2 مستخدما بيانات المثال (١). $(n = 10, s = 4)$.

لدينا:

$$s^2 = 16 \qquad \chi^2(.01;9) = 2.09 \qquad \chi^2(.99;9) = 21.67$$

$$6.6 = \frac{9(16)}{21.67} \le \sigma^2 \le \frac{9(16)}{2.09} = 68.9$$

اختبارات

يمكن وضع اختبارات وحيدة الجانب وثنائية الجانب حول تباين المجتمع σ^2 باستخدام (1.67). ويحوي الجدول (١–٣) قواعد القرار لكل من الحالات الثلاث الممكنة مع إبقاء مخاطرة التورط بخطأ من النوع الأول عند المستوى α.

(١–٩) مقارنات تباين مجتمعين - مجتمعات طبيعية

اختيرت عينتان مستقلتان من مجتمعين طبيعيين بمتوسطين وتباينين μ_1, σ_1^2 و μ_2, σ_2^2، على الترتيب. وباستخدام رموز الفقرة (١–٧) فإن تبايني العينتين هما:

$$s_1^2 = \frac{\sum_i (Y_i - \overline{Y})^2}{n_1 - 1} \tag{1.69a}$$

$$s_1^2 = \frac{\sum_i (Z_i - \overline{Z})^2}{n_2 - 1} \tag{1.69b}$$

جدول (١ - ٣). قواعد القرار لاختبارات حول تباين مجتمع طبيعي σ^2

البدائل	قاعدة القرار
	(أ)
$H_0:\sigma^2=\sigma_0^2$ $H_a:\sigma^2\neq\sigma_0^2$	إذا كان $\chi^2(\alpha/2;n-1)\leq\frac{(n-1)s^2}{\sigma_0^2}\leq\chi^2(1-\alpha/2;n-1)$ استنتج H_0 وفيما عدا ذلك استنتج H_a
	(ب)
$H_0:\sigma^2\geq\sigma_0^2$ $H_a:\sigma^2<\sigma_0^2$	إذا كان $\frac{(n-1)s^2}{\sigma_0^2}\geq\chi^2(\alpha;n-1)$ استنتج H_0 إذا كان $\frac{(n-1)s^2}{\sigma_0^2}<\chi^2(\alpha;n-1)$ استنتج H_a
	(ج)
$H_0:\sigma^2\leq\sigma_0^2$ $H_a:\sigma^2>\sigma_0^2$	إذا كان $\frac{(n-1)s^2}{\sigma_0^2}\leq\chi^2(1-\alpha;n-1)$ استنتج H_0 إذا كان $\frac{(n-1)s^2}{\sigma_0^2}>\chi^2(1-\alpha;n-1)$ استنتج

ولدينا:

$$\frac{s_1^2}{\sigma_1^2}\div\frac{s_2^2}{\sigma_2^2} \tag{1.70}$$

يتوزع وفق $F(n_1-1\ ,\ n_2-1)$ عندما تكون العينتان المستقلتان من مجتمعين طبيعيين.

تقدير فترة

نحصل على حد ثقة أدنى L وحد ثقة أعلى U لـ $\frac{\sigma_1^2}{\sigma_2^2}$ بمعامل ثقة $1-\alpha$ من استخدام العلاقة (1.70):

$$L = \frac{s_1^2}{s_2^2}\left[\frac{1}{F(1-\alpha/2; n_1-1, n_2-1)}\right]$$
$$U = \frac{s_1^2}{s_2^2}\left[\frac{1}{F(\alpha/2; n_1-1, n_2-1)}\right] \tag{1.71}$$

جدول (١-٤). قواعد القرار لاختبارات حول تباينين σ_1^2 , σ_2^2 لمجتمعين طبيعيي - عينات مستقلة

قاعدة القرار	البدائل
(أ)	
إذا كان $F(\alpha/2; n_1-1, n_2-1) \le \frac{s_1^2}{s_2^2} \le F(1-\alpha/2; n_1-1, n_2-1)$ استنتج H_0 وفيما عدا ذلك استنتج H_a	$H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \ne \sigma_2^2$
(ب)	
إذا كان $\frac{s_1^2}{s_2^2} \ge F(\alpha; n_1-1, n_2-1)$ استنتج H_0 إذا كان $\frac{s_1^2}{s_2^2} < F(\alpha; n_1-1, n_2-1)$ استنتج H_a	$H_0: \sigma_1^2 \ge \sigma_2^2$ $H_a: \sigma_1^2 < \sigma_2^2$
(ج)	
إذا كان $\frac{s_1^2}{s_2^2} \le F(1-\alpha; n_1-1, n_2-1)$ استنتج H_0 إذا كان $\frac{s_1^2}{s_2^2} > F(1-\alpha; 2; n_1-1, n_2-1)$ استنتج H_a	$H_0: \sigma_1^2 \le \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$

مثال (٩). أوجد 90 بالمائة فترة ثقة لـ σ_1^2/σ_2^2 في حالة البيانات التالية:

$$n_1 = 16 \qquad n_2 = 21$$
$$s_1^2 = 54.2 \qquad s_2^2 = 17.8$$

لدينا:

$$F(.05;15,20) = 1/F(.95;20,15) = 1/2.33 = .429$$
$$F(.95;15,20) = 2.20$$

$$1.4=\frac{54.2}{17.8}\left(\frac{1}{2.20}\right)\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{54.2}{17.8}\left(\frac{1}{.429}\right)=7.1$$

اختبارات

يمكن وضع اختبارات وحيدة الجانب وثنائية الجانب حول σ_1^2/σ_2^2 بالاستفادة من (1.70). ويحوي الجدول (١-٤) قواعد القرار لكل من الحالات الثلاث الممكنة مع إبقاء مخاطرة التورط بخطأ من النوع الأول عند α.

مثال (١٠). اختر بين البدائل:

$$H_0:\sigma_1^2=\sigma_2^2$$
$$H_a:\sigma_1^2\ne\sigma_2^2$$

مقيدا α عند 0.02 ومستخدما بيانات المثال (٩).

لدينا:

$$F(.01;15,20)=1/F(.99;20,15)=1/3.37=.297$$
$$F(.99;15,20)=3.09$$

وبالتالي فإن قاعدة القرار هي: إذا كانت $.297\le\frac{s_1^2}{s_2^2}\le3.09$ استنتج H_0، خلاف ذلك قرر H_a.

وحيث إن $\frac{s_1^2}{s_2^2}=54.2/17.8=3.04$ فإننا نستنتج H_0.

الباب الأول

الانحدار الخطي البسيط

- الانحدار الخطي بمتغير مستقل واحد
- استقراءات في تحليل الانحدار
- تشخيصات وتدابير علاجية
- استقراءات متزامنة ومواضيع أخرى في تحليل الانحدار
- طريقة المصفوفة في تحليل انحدار خطي بسيط

الانحدار الخطي بمتغير مستقل واحد

تحليل الانحدار هو أداة إحصائية تستفيد من العلاقة بين متغيرين كميين أو أكثر للتنبؤ بأحد المتغيرات استنادا إلى قيم المتغير أو المتغيرات الأخرى. فمثلا إذا علمنا العلاقة بين مصروفات الدعاية وبين المبيعات، فيمكننا الاستفادة من تحليل الانحدار للتنبؤ بالمبيعات حالما تتوفر لنا قيمة نفقات الدعاية.

في القسم الأول من هذا الكتاب نتبنى تحليل الانحدار عند استخدام متغير واحد للتنبؤ بالمتغير قيد الاهتمام. وفي هذا الفصل بالتحديد، سوف نتناول الأفكار الأساسية لتحليل الانحدار ونناقش تقدير معالم نموذج الانحدار.

(٢-١) العلاقات بين المتغيرات.

مفهوم العلاقة بين متغيرين، مثل نفقات السكن لأسرة ودخل الأسرة هو مفهوم مألوف. وسوف نميز بين *علاقة دالّية* و*علاقة إحصائية* ونتعرض لكل منهما بدورها.

العلاقات الدالّية بين متغيرين

يعبر عن العلاقة الدالّية بين متغيرين بصيغة رياضية. فإذا كان X المتغير المستقل، وY المتغير التابع، فإن العلاقة الدالية تكون من الشكل:

$$Y = f(X)$$

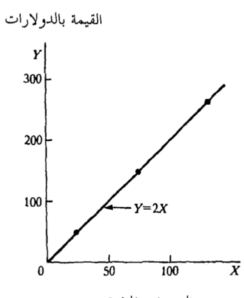
وإذا أُعطيت قيمة معينة لـ X فإن الدالّة f تشير إلى قيمة Y المقابلة.

مثال. اعتبر العلاقة بين قيمة المبيعات بالدولارات (Y) لسلعة تباع بسعر محدد وعدد الوحدات المباعة (X). إذا كان سعر البيع $2 للوحدة، فيعبر عن العلاقة بين X و Y بالمعادلة:

$$Y = 2X$$

ويمثل الشكل (٢-١) هذه العلاقة الدالّية.

شكل (٢-١) مثال لعلاقة دالية



فيما يلي عدد الوحدات المباعة والمبيعات بالدولار خلال الفترات الثلاث السابقة (حيث بقي سعر الوحدة ثابتا عند $2):

الفترة	عدد الوحدات المباعة	القيمة بالدولارات
1	75	$150
2	25	$50
3	130	$260

مُثّلت هذه القيم أيضا في الشكل (٢-١). لاحظ أنها جميعا وقعت مباشرة على خط العلاقة الدالّية. وهذه خاصة مميزة لجميع العلاقات الدالّية.

العلاقة الإحصائية بين متغيرين

على عكس العلاقة الدالّية فالعلاقة الإحصائية ليست تماما دالّة. وبوجه عام، لاتقع مشاهدات الدالّة الاحصائية تماما على منحنى العلاقة بينهما.

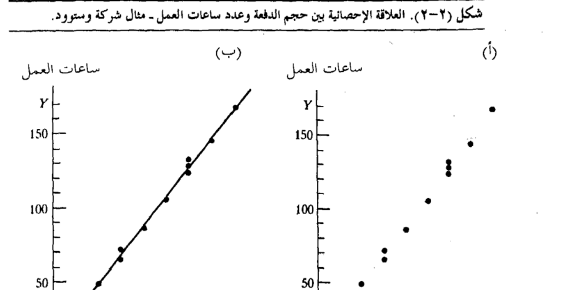
مثال(١). تصنع شركة ويستوود قطع غيار معينة على شكل دفعات شهرية. وتختلف في حجمها تبعا لتذبذب الطلبات ويحوي الجدول (٢-١) صفحة (٤٦) بيانات عن حجم الدفعات وعدد ساعات العمل لعشر دفعات متتابعة حديثة أُنتجت تحت شروط إنتاج متشابهة. وقد رُسمت هذه البيانات في الشكل(٢-٢) أ حيث اتُّخذت ساعات العمل كمتغير تابع أو استجابة Y واتُّخذ حجم الدفعة X متغيرا مستقلا أو متغير تنبؤ. وقد تمَّ الرسم كما في السابق. فعلى سبيل المثال، رُسمت نتيجة أول دفعة إنتاج بحيث كان $X=30$ و $Y=73$.

من الواضح، أن الشكل (٢-٢) أ يقترح وجود علاقة بين حجم الدفعة وعدد ساعات العمل، بمعنى أنه كلما ازداد حجم الدفعة اتجه عدد ساعات العمل اللازمة إلى الازدياد. وعلى كل حال، فالعلاقة ليست علاقة تامة. ويوحي وجود نقاط مبعثرة، بأن بعض الانحرافات في عدد ساعات العمل غير مؤثر في حجم الدفعة. فمثلا عدد الوحدات المنتجة في كل من دفعتي الإنتاج (1 و 8) هي 30 قطعة بينما تتطلبان إلى حد ما ساعات عمل مختلفة. ونظرا لانتشار النقاط في العلاقة الإحصائية فإن الشكل (٢- ٢)أ يسمى مخطط انتشار أو رسم انتشار، وبالمصطلح الإحصائي تسمى كل نقطة في شكل الانتشار محاولة (تكرار) أو حالة.

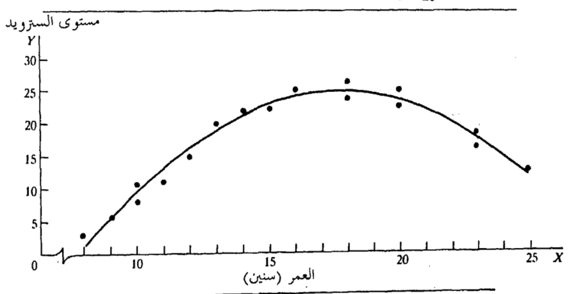
في الشكل (٢-٢) ب رسمنا خطا لعلاقة تصف العلاقة الإحصائية بين ساعات العمل وحجم الدفعة. وتشير هذه العلاقة إلى النزعة العامة التي تتغير بموجبها ساعات العمل مع تغير حجم الدفعة. ويُلاحظ هنا عدم وقوع معظم النقاط على خط العلاقة الإحصائية تماما. ويمثل تبعثر النقاط حول الخط التغير في عدد ساعات العمل الذي لايُعزى إلى تغير حجم الدفعة وإنما يعتبر عادة من طبيعة عشوائية. ومع أن العلاقات الإحصائية ليست في دقة العلاقة الدالّية إلا أنها ذات فوائد جمة.

مثال(٢). يمثل الشكل (٢-٣) بيانات أعمار ومستوى الستروئيد في البلازما لـ 17 من الإناث المتصفات بالسمنة واللآتي تتراوح أعمارهن بين 18 و 25 سنة. وتقترح البيانات بقوة أن العلاقة الإحصائية هي علاقة منحنية (ليست خطية). وقد رُسم منحنى العلاقة الإحصائية في الشكل (٢-٣). ويتضح من المنحنى أنه كلما زاد العمر فإن مستوى الستروئيد يزداد إلى أن يصل نقطة يبدأ بعدها بالهبوط.ونلاحظ أن تبعثر النقاط حول منحني العلاقة الإحصائية هي سمة عامة لكل العلاقات الإحصائية.

شكل (٢-٢). العلاقة الإحصائية بين حجم الدفعة وعدد ساعات العمل - مثال شركة وسترود.



شكل (٢-٣) علاقة إحصائية منحنية بين العمر ومستوى الستروید في الإنـاث السـمینات بعمر 18 إلى 25 سنة.



(٢–٢) نماذج انحدار واستخداماتها

الأصول التاريخية

إن أول من طوَّر تحليل الانحدار هو السير فرانسيس كالتون (*F. Galton*) في الجزء الأخير من القرن التاسع عشر. فقد درس كالتون العلاقة بين أطوال الآباء والأبناء، ولاحظ أن أطوال أبناء لآباء طوال أو قصار تبدو وكأنها "ترتد" أو "تنحدر" نحو متوسط المجموعة. واعتبر كالتون هذه النزعة رجعة إلى التوسط. وقد طوّر كالتون وصفا رياضيا لهذه النزعة التراجعية يشكل الأصل التاريخي لما يُعرف اليوم بنماذج الانحدار. ويستمر مصطلح الانحدار إلى يومنا هذا كوصف لعلاقات إحصائية بين متغيرات.

مفاهيم أساسية

إن نموذج الانحدار ماهو إلا وسيلة رسمية للتعبير عن عنصرين أساسيين من عناصر العلاقة الإحصائية.

١ـ نزوع المتغير التابع Y للتغير مع المتغير المستقل X بصورة نمطية.

٢ـ تبعثر النقاط حول منحني العلاقة الإحصائية.

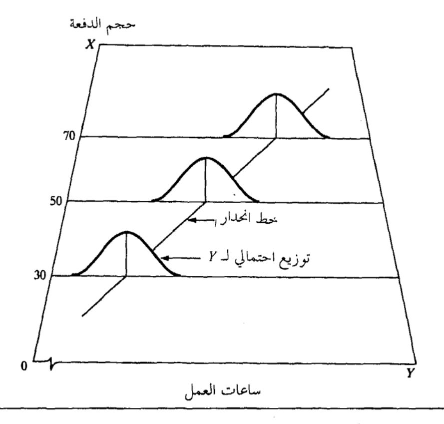
وقد تجسّدت هاتان الخاصتان في نموذج الانحدار من خلال الافتراضين التاليين:

١ـ يوجد توزيع احتمالي للمتغير Y عند كل مستو من مستويات X.

٢ـ تتغير متوسطات هذه التوزيعات الاحتمالية بصورة نظامية مع تغيرX.

مثال. اعتبر مرة أخرى مثال حجم الدفعات لشركة وستوود. في نموذج الانحدار، نعتبر عدد ساعات العمل كمتغير عشوائي. ولكل حجم دفعة، نفترض توزيعا احتماليا Y. ويبين الشكل (٢ـ٤) مثل ذلك التوزيع الاحتمالي المقابل لـ X=30. وهو حجم الدفعة الأولى في الجدول (٢ـ١). وننظر هنا إلى قيمة Y الفعلية وهي 73 في مثالنا في الجدول (٢ـ١) على أنها اختيار عشوائي من هذا التوزيع الاحتمالي.

شكل (٢ ـ ٤) تمثيل تصويري لنموذج انحدار خطي



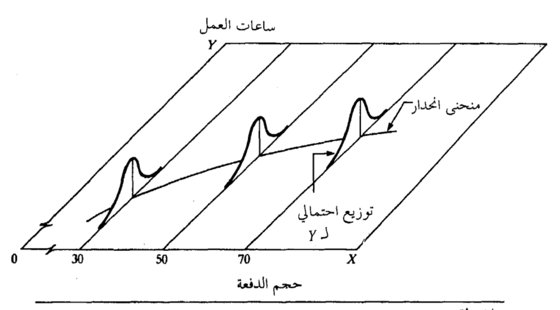
ويوضح الشكل (٢ـ٤) أيضا توزيعين احتماليين لِـ Y يقابلان حجمي الدفعتين X=50 و X=70. لاحظ أن متوسطات هذه التوزيعات الاحتمالية لها علاقة نظامية مع مستوى X. وتسمى هذه العلاقة النظامية *دالّة انحدار* Y على X. ويسمى رسم دالّة الانحدار منحني *الانحدار*. لاحظ أن منحني الانحدار خطّي في الشكل (٢ـ٤). وهذا بدوره يؤدي في مثالنا إلى أن توقع (متوسط) عدد ساعات العمل يتغير خطيا مع حجم الدفعة.

ولايوجد بالطبع أي سبب مُسبق يستوجب ارتباط عدد ساعات العمل خطيا مع حجم الدفعة. ويوضح الشكل (٢–٥) نموذج انحدار آخر لمثالنا، دالّة الانحدار فيه منحنية، ومع ازدياد حجم الدفعة X، يعكس الشكل قيما لِـY أقل مما لو كانت العلاقة خطية. ويختلف الشكل (٢–٥) في توجهه عن الشكل (٢–٤) من حيث إن المحور X

والمحور Y قد رُسما بالطريقة التقليدية في الشكل (٢-٥)، وبينما لا يسمح هذا بوضوح لمنظر التوزيعات الاحتمالية كوضوحه في الشكل (٢-٤)، إلا أن توجّه الشكل (٢-٥) يبين منحنى الانحدار بمنظور واضح وهو ماسنستخدمه من الآن فصاعدا.

وقد تختلف نماذج الانحدار إما في شكل دالّة الانحدار، كما هو الحال في الشكلين (٢-٤) و(٢-٥)، أو في شكل توزيع Y الاحتمالي أو بطرق أخرى. ومهما كان الاختلاف يبقى توزيع Y الاحتمالي عند قيمة معطاة لـ X المقابل الرياضي لظاهرة التغير التجريبي في علاقة إحصائية. وبالمثل فإن منحنى الانحدار، الذي يصف العلاقة بين متوسطات التوزيعات الاحتمالية وبين X هو، في علاقة إحصائية، التعبير الرياضي عن نزوع قيم Y إلى التغير مع قيم X بصورة نظامية.

شكل (٢-٥) تمثيل تصويري لنموذج انحدار منحني



ملاحظة

التعبير عن X على أنه " **متغير مستقل**" أو **متغير تنبؤ**"، والتعبير عن Y على أنه " **متغير تابع**" أو"**متغير استجابة** "هي ألقاب اصطلاحية. ولاتتضمن، في حالة معينة، أن Y تعمتد اعتمادا سببيا على X. وبصرف النظر عن قوة العلاقة الإحصائية فإن

النموذج الإحصائي لاينطوي بالضرورة على ثنائية السبب والنتيجة. وفي بعض التطبيقات يعتمد المتغير المستقل اعتمادا سببيا على متغير الاستجابة، كما هو الحال عندما نقدر درجة الحرارة (الاستجابة) من ارتفاع الزئبق (المتغير المستقل) في ميزان الحرارة.

نماذج انحدار بأكثر من متغير مستقل واحد. قد تحتوي نماذج الاحتمال على أكثر من متغير مستقل واحد.

١- في تطبيق لنموذج تحليل انحدار تناول 67 مكتبا فرعيا من مكاتب "السلسلة المالية للمستهلك" اتّخذت " تكلفة التشغيل المباشرة " للسنة المنتهية كمتغير استجابة. ويوجد أربعة متغيرات مستقلة هي متوسط حجم القروض غير المدفوعة خلال السنة، ومتوسط عدد القروض غير المدفوعة، والعدد الكلي لطلبات القروض المقدمة، والمؤشر القياسي لسلّم الرواتب.

٢- في دراسة تناولت شراء محاريث آلية كان حجم المحاريث المشتراة (مقاسا بعدد الأحصنة)، من منطقة بيع لشركة للمعدات الزراعية، هو متغير الاستجابة. وهناك تسعة متغيرات مستقلة تتضمن متوسط عمر المحاريث في مزارع المنطقة، وعدد المزارع في المنطقة، ومؤشر كمي لإنتاج المحاصيل في المنطقة.

٣- في دراسة طبية عن قصار الأطفال، كان مستوى الذروة لبلازما هرمون النمو هو متغير الاستجابة. ويوجد أربعة عشر متغيرا مستقلا تتضمن العمر والجنس والطول والوزن وعشرة قياسات تحت الجلد.

عندما يكون هناك أكثر من متغير مستقل، فينبغي تعميم الخواص الممثلة في الشكلين (٢-٤) و(٢-٥) إلى عدد أكبر من الأبعاد. فعلى سبيل المثال، يفترض نموذج الانحدار وجود توزيع احتمالي للمتغير لِـ Y من أجل كل مركب (X_1, X_2) من مستويات المتغيرين المستقلين، ويمثل سطح الانحدار عندئذ، العلاقة النظامية بين متوسطات هذه التوزيعات الاحتمالية والمتغيرين المستقلين X_1 و X_2.

بناء نماذج الانحدار

اختيار المتغيرات المستقلة. بما أنه ينبغي لنا، عند بنائنا النماذج، اختزال الواقع الفعلي إلى جزء طيّع منه، يمكن التعامل معه، فينبغي أن يقتصر نموذج الانحدار لأي

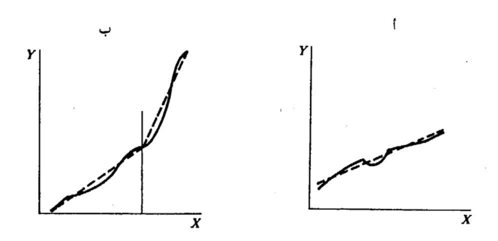
مسألة ندرسها على عدد محدود من المتغيرات المستقلة. ولذلك فإن المشكلة الأساسية تكمن في اختيار مجموعة من المتغيرات المستقلة لنموذج الانحدار يمكن أن توصف، بمعنى ما، أنها، ولأغراض التحليل، متغيرات جيدة. والعامل الرئيس في اختيار متغير مستقل هو مدى مساهمته في تخفيض ماتبقى من التغير في Y، بعد أن تكون مساهمات متغيرات أخرى، تمَّ مبدئيا اختيارها إلى النموذج، قد أُخذت في الاعتبار. ومن العوامل الأخرى تأتي أهمية المتغير كعامل سببي في العملية موضع التحليل. ودرجة الدقة، وسرعة الحصول على مشاهدات المتغير، وتكلفتها، مقارنة بمتغيرات أخرى منافسة. وكذلك إمكانية وضع المتغير تحت إدارة المجرب. وسوف نناقش في الفصل الثاني عشر أساليب ومشاكل اختيار المتغيرات المستقلة التي يتضمنها نموذج انحدار.

الشكل الدالي لعلاقة انحدار. يرتبط اختيار الشكل الدالي لمعادلة الانحدار باختيار المتغيرات المستقلة. ففي بعض الأحيان يمكن أن تشير المعرفة النظرية المتيسرة إلى الشكل الدالي. فعلى سبيل المثال، قد تشير نظرية التعلّم إلى أن دالّة الانحدار التي تربط تكلفة وحدة إنتاج بعدد مرات إنتاجها سابقا، ينبغي لها أن تتخذ شكلا محددا بخواص مقاربة معينة.

وفي الغالب ـ على كل حال ـ لايكون الشكل الدالّي لعلاقة الانحدار معروفا سلفا، ولابد من اتخاذ قرار بشأنها حالما يتم جمع البيانات وتحليلها. وهكذا استُخدمت، في الغالب، علاقات انحدار خطية وتربيعية كتقريبات أولية مُرضية لدوال انحدار من طبيعة غير معروفة. وفي الحقيقة يمكن استخدام هذه الأنواع البسيطة من دوال الانحدار حتى لو كانت النظرية تجهّزنا بالشكل الدالّي المناسب، خاصة عندما يكون الشكل معقدا جدا، ولكن يمكن تقريبه بشكل معقول بدالّة انحدار خطية أو تربيعية. ويوضح الشكل (٢-٦) أ حالة دالّة انحدار معقدة وتقريبها بانحدار خطي معقول. ويقدم الشكل (٢- ٦) ب مثالا لإمكانية استخدام دالتي انحدار خطيتين، واحدة تلو الأخرى، لتقريب دالّة انحدار معقدة.

مجال النموذج. عند صياغة نموذج انحدار، نحتاج عادة إلى تقييد تغطية النموذج بحيث تقتصر على فترة أو منطقة من القيم للمتغير أو المتغيرات المستقلة. ويتحدد المجال من خلال تصميم الدراسة أو من خلال مدى البيانات المتوفرة. فمثلا، قد تدرس شركة تأثير السعر على حجم المبيعات من خلال ستة تسعيرات تبدأ من $4.95 إلى $6.95 ويتحدد مجال النموذج هنا بين مايقارب $5 إلى مايقارب $7. وقد يعاني شكل دالّة الانحدار من شكّ كبير خارج هذا المدى على وجه الخصوص ذلك لأن الدراسة لاتقدم أية دلائل عن طبيعة العلاقة الإحصائية تحت $4.95 أوفوق $6.95.

شكل (٢ ـ ٦). استخدام دوال الانحدار الخطي لتقريب دوال انحدار معقدة



استخدامات تحليل الانحدار

يخدم تحليل الانحدار ثلاثة أغراض رئيسة (١) الوصف ، (٢) السيطرةو (٣) التنبؤ، كما وُضّح في الأمثلة الثلاثة المذكورة سابقا. فمثال شراء الحراثات يخدم الوصف. وتخدم دراسة تكاليف تشغيل المكاتب الفرعية هدف السيطرة الإدارية، حيث استطاعت الإدارة من خلال تطوير علاقات إحصائية بين التكاليف والمتغيرات المستقلة في النظام، وضع معايير للتكلفة لكل مكتب فرعي من مكاتب الشركة. وكان التنبؤ هو الهدف من الدراسة الطبية للأطفال القصار إذ استطاع الإكلينيكيون استخدام

العلاقة الإحصائية للتنبؤ بنقص هرمون النمو في الأطفال القصار مستخدمين قياسات بسيطة تمت على الأطفال.

وتتداخل الأغراض المتعددة لتحليل الانحدار في الواقع العملي. ويجهزنا مثال حجم الدفعة لشركة وستوود بأحد الحالات. إن معرفة العلاقة بين حجم الدفعة وبين عدد ساعات العمل في فترات إنتاج سابقة تسمح للإدارة بالتنبؤ بساعات العمل اللازمة لفترة إنتاج مقبلة يُعرف فيها حجم الدفعة، وذلك لأغراض تقدير الكلفة وجدولة الانتاج. وبعد إكمال دورة الانتاج تستطيع الإدارة مقارنة ساعات العمل الحقيقية بالساعات المتنبأ بها لفرض السيطرة الإدارية.

(٢-٣) نموذج انحدار بتوزيع غير معروف لحد الخطأ

عبارة رسمية للنموذج

في القسم الأول من هذا الكتاب، نعتبر نموذج انحدار أساسي بمتغير مستقل واحد و دالة انحدار خطية. ويمكن عرض النموذج كما يلي:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (2.1)$$

حيث:

Y_i قيمة متغير الاستجابة في التكرار i

β_0 و β_1 معلمتان

X_i ثابت معلوم، ونقصد، قيمة المتغير المستقل في التكرار i

ε_i حد خطأ عشوائي متوسطه $E(\varepsilon_i)$ وتباينه $\sigma^2(\varepsilon_i) = \sigma^2$ و ε_i غير مرتبطين.

أي أن $cov\{\varepsilon_i, \varepsilon_j\} = 0$ لكل $i \neq j$ ، $i = 1, \ldots, n$

يقال عن *نموذج الانحدار* (2.1) أنه بسيط، وخطي في المعالم وخطي في المتغير المستقل. فهو بسيط لأنه يستخدم متغيرا مستقلا واحدا فقط، و"خطي المعالم" لأنه لاتظهر أي معلمة كأس أو مضروبة بمعلمة أخرى أو مقسومة على معلمة أخرى، و"خطي في المتغير المستقل" لأن هذا المتغير لايظهر إلا مرفوعا للأس واحد. ويسمى النموذج الخطي في المعالم والخطي في المتغير المستقل نموذجا من المرتبة الأولى أيضا.

سمات مهمة للنموذج

١– القيمة المشاهدة لِـ Y في التجربة i هي مجموع مركبتين (١) الحد الثابت. $\beta_0 + \beta_1 X_i$ و(٢) الحد العشوائي ε_i. وبالتالي يكون Y_i متغيرا عشوائيا.

٢– بما أن $E(\varepsilon_i) = 0$ فإنه يتبع من (1.13c) أن:

$$E\{Y_i\} = E\{\beta_0 + \beta_1 X_i + \varepsilon_i\} = \beta_0 + \beta_1 X_i + E\{\varepsilon_i\} = \beta_0 + \beta_1 X_i$$

لاحظ أن $\beta_0 + \beta_1 X_i$ يلعب دور الثابت a في النظرية (1.13c).

وهكذا تأتي الاستجابة Y_i، عندما يكون مستوى X في التجربة i هو X_i، من توزيع احتمالي توقعه:

$$E\{Y_i\} = \beta_0 + \beta_1 X_i \qquad (2.2)$$

ومن ثم نعلم أن دالّة الانحدار للنموذج (2.1) هي:

$$E\{Y\} = \beta_0 + \beta_1 X \qquad (2.3)$$

ذلك لأن دالّة الانحدار تربط متوسطات توزيعات Y الاحتمالية الموافقة لقيم معطاة لِـ X بمستوى X.

٣ ـ تتجاوز قيمة Y التي شوهدت في التجربة i قيمة دالّة الانحدار، أو تقلّ عنها، بحدّ خطأ قدره ε_i.

٤– يُفترض لحد الخطأ ε_i تباين ثابت σ^2. وينتج عن ذلك أن للاستجابات التباين الثابت نفسه:

$$\sigma^2\{Y_i\} = \sigma^2 \qquad (2.4)$$

فمن النظرية (1.16a) لدينا:

$$\sigma^2\{\beta_0 + \beta_1 X_i + \varepsilon_i\} = \sigma^2\{\varepsilon_i\} = \sigma^2$$

وهكذا يفترض نموذج الانحدار (2.1) أن لتوزيعات Y الاحتمالية التباين σ نفسه، وذلك بغض النظر عن مستوى المتغير المستقل X.

٥– يُفترض أن حدود الخطأ غير مرتبطة. لذا لاتؤثر نتيجة أي تكرار للتجربة على حد الخطأ لتكرار آخر، من حيث كونه موجبا أم سالبا، صغيرا أم كبيرا. وبما أن حدي الخطأ ε_i و ε_j غير مرتبطين فإن الاستجابتين Y_i و Y_j غير مرتبطتين.

٦– والخلاصة، يتضمن نموذج الانحدار (2.1) أن مشاهدات متغير الاستجابة تأتي من توزيعات احتمالية توقعاتها $E(Y_i) = \beta_0 + \beta_1 X_i$، وتبايناتها σ^2 هي نفسها لكل مستويات X، بالإضافة إلى كون أي مشاهدتين Y_i و Y_j غير مرتبطتين.

مثال.

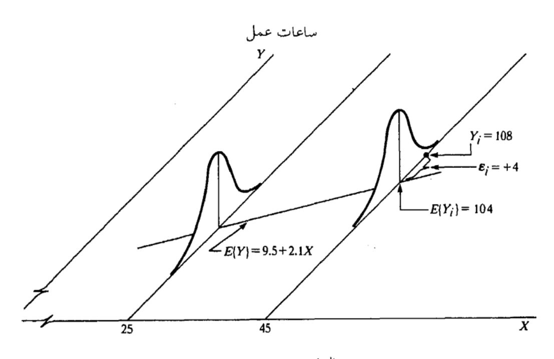
لنفترض أن نموذج الانحـدار (2.1) قـابل للتطبيـق علـى حجـم الدفعـة في مثـال شركة وستوود وهو كما يلي:

$$Y_i = 9.5 + 2.1X_i + \varepsilon_i$$

ويتضمن الشكل (٢-٧) تمثيلا لدالّة الانحدار:

$$E(Y) = 9.5 + 2.1X$$

شكل (٢-٧). توضيح لنموذج الانحدار الخطي (2.1)



لنفرض أن التكرار i أنتج دفعة حجمها $X_i = 45$ وحدة، وكان عدد سـاعات العمـل الحقيقية $Y_i = 108$. ففي هذه الحالة تكون قيمة حد الخطأ $\varepsilon_i = +4$، ذلك لأن:

$$E\{Y_i\} = 9.5 + 2.1(45) = 104$$

و:

$$Y_i = 108 = 104 + 4$$

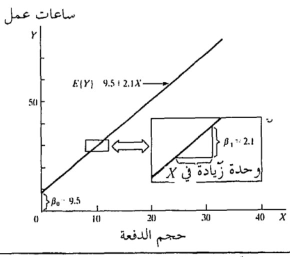
ويعرض الشكل (٢-٧) توزيـع Y الاحتمـالي عندمـا $X = 45$ ، ويشـير إلى الموقع من هذا التوزيع الذي جاءت منه المشـاهدة $Y_i = 108$. ونلاحـظ ثانيـة أن حـد الخطأ هو ببساطة انحراف Y_i عن قيمته المتوسطة $E\{Y_i\}$.

ويعرض الشكل (٢-٧) أيضا توزيع Y الاحتمـالي عندمـا $X = 25$. لاحـظ أن هذا التوزيع يُظهر نفس تشتت التوزيـع الاحتمـالي عندمـا $X = 45$، وذلـك امتثـالا لمتطلبات نموذج الانحدار (2.1).

معاني معالم الانحدار

تسمى المعالم β_0 و β_1 في نموذج الانحدار (2.1) ***معاملات الانحدار***. وتشـير β_1، وهي ميل خط الانحدار إلى التغير في متوسط توزيع Y الاحتمالي لكل وحدة زيادة في X. والمعلمة β_0 هي التقاطع الصادي لخط الانحدار. إذا تضمن مدى النموذج القيمة $X = 0$، فإن β_0 تعطي متوسط توزيع Y الاحتمالي عنـد $X=0$. وليس للمعلمة β_0 أي تفسير خاص بها كحد منفصل في نموذج الانحدار إذا لم يتضمن مجاله القيمة $X=0$.

شكل (٢-٨) معنى معلمتي نموذج انحدار خطي بسيط



مثال. يعرض الشكل (٢-٨) دالّة الانحدار:

$$E\{Y\} = 9.5 + 2.1X$$

لحجم الدفعة في مثال شركة وستوود السابق. ويشير الميل $\beta_1 = 2.1$ إلى أن زيادة وحـدة واحدة في حجم الدفعة يقود إلى زيادة 2.1 ساعة عمل في متوسط توزيع Y الاحتمالي.

ويشير الجزء المقطوع $\beta_0 = 9.5$ إلى قيمة دالّة الانحدار عند $X = 0$. ومع ذلك فإن β_0لا تحمل لذاتها معنى جوهريا باعتبار أن نموذج الانحدار الخطي قد وُضع ليُطبق على دفعات تتراوح حجومها بين 20 إلى 80 وحدة. وعلى وجه الخصوص فإنها لاتشير بالضرورة إلى المتوسط الزمني عند بدء العملية (متوسط عدد ساعات العمل قبل بداية الإنتاج الفعلي). ولربما تطلّبَ الأمر استخدام نموذج انحدار منحنٍ، مع قيمة لِ β_0مختلفة عن القيمة المستخدمة في النموذج الخطي، إذا ماأريد تمديد مجال النموذج إلى دفعات يصل حجمها إلى الصفر.

صور بديلة لنموذج الانحدار

من الملائم في بعض الأحيان كتابة (2.1) بأشكال مختلفة بعض الشيء ولكنها متكافئة. ليكن X_0 متغير دمية مطابقا ومساويا للواحد وعندئذ يمكن كتابة (2.1) كما يلي:

$$Y_i = \beta_0 X_0 + \beta_1 X_i + \varepsilon_i \qquad \text{حيث } X_0 \equiv 1 \tag{2.5}$$

واستخدام الانحراف $X_i - \overline{X}$ كمتغير مستقل بدلا من X_i يعطي بديلا آخر مفيدا في بعض الأحيان. ولإبقاء (2.1) على ما هو عليه، نحتاج إلى كتابة:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 (X_i - \overline{X}) + \beta_1 \overline{X} + \varepsilon_i \\ &= (\beta_0 + \beta_1 \overline{X}) + \beta_1 (X_i - \overline{X}) + \varepsilon_i \\ &= \beta_0^* + \beta_1 (X_i - \overline{X}) + \varepsilon_i \end{aligned}$$

وهكذا تكون النسخة البديلة للنموذج:

$$Y_i = \beta_0^* + \beta_1 (X_i - \overline{X}) + \varepsilon_i \tag{2.6}$$

حيث:

$$\beta_0^* = \beta_0 + \beta_1 \overline{X} \tag{2.6a}$$

وسنستخدم النماذج (2.1) ، (2.5) و (2.6) بالتبادل وفق ما تمليه المناسبة.

(٢-٤) بيانات تحليل الانحدار

عادة، لا نعرف قيم معالم الانحدار β_0و β_1في نموذج الانحدار (2.1). ونحتاج إلى تقديرهما من بيانات مناسبة. وفي الحقيقة، وكما ذكرنا سابقا، ليس لدينا في

معظم الأحيان، معرفة مسبقة وكافيـة عـن المتغـيرات المسـتقلة المناسـبة، وعـن الشـكل الدالّي لعلاقة الانحدار، (مثلا،خطية أو منحنية) ونحتاج إلى الاعتماد على خواص مميزة للبيانات كي نطور نموذج الانحدار المناسب.

ويمكن الحصـول علـى بيانـات لتحليـل الانحـدار بطـرق غـير تجريبيـة وتجريبيـة وسوف نتطرق لكل منهما بدوره.

بيانات المشاهدة

بيانات المشاهدة هي بيانات غير تجريبية. حيث نحصـل علـى بيانـات مـن هـذا القبيل بدون السـيطرة علـى المتغـير (المتغـيرات) المسـتقل (المسـتقلة) موضـع الاهتمـام. فمثلا، رغب مسؤولو شركة دراسة العلاقة بين عمر المستخدم X وعدد أيام المـرض Y خلال السنة الماضية. فاستخدموا لتحليل الانحدار بيانات تم الحصول عليها من الملفـات الشخصية، وبيانات كهـذه هـي بيانـات مشـاهدة. إذ لا يمكـن السـيطرة علـى المتغـير المستقل.

وقد اعتمدت شركة وستوود في مثال حجم الدفعة المذكور سابقا على بيانات تاريخية. وهذه أيضا بيانات مشاهدة، حيث يتحدد حجم الدفعات وفقـا للطلـب علـى المنتج و لم تتحكم به شروط تجريبية.

وفي معظم الأحيان يستند تحليل الانحدار على بيانات مشاهدة لأنه لا يمكننـا في الغالب، القيام بتجارب نتحكم في مسارها. وفي مثال منسـوبي الشـركة، علـى سـبيل المثال، ليس بالإمكان السيطرة على العمر بتخصيص أعمار للأشخاص.

ومن العيوب الرئيسة لبيانـات المشـاهدة أنهـا لا تزودنـا بمعلومـات كافيـة عـن العلاقات بين السبب والتأثير. فمثلا قد لا تعـني العلاقـة الإيجابيـة بـين عمـر المسـتخدم وعدد أيام المرض في مثال أفراد الشركة أن عدد أيام المرض هو النتـاج المباشـر للعمـر. فلربمـا كـان مسـتخدمو الشـركة الشـباب يعملـون أساسـا في الخـارج في حـين يعمـل المستخدمون الكبار، عادة، في الداخل، ويتحمل مكان العمل المسؤولية الأهم في عـدد أيام المرض.

وحينما يعتمد تحليل انحدار، نقوم به لأغراض وصفية، على بيانات مشاهدة فإنه ينبغي تقصّي ما إذا كانت هناك متغيرات مستقلة غير المتغيرات المعتمدة في النموذج، يمكنها أن تفسر بصورة أفضل علاقات سبب وتأثير.

بيانات تجريبية

وكثيرا ما يمكن إجراء تجربة نتحكم فيها لتزودنا ببيانات تمكّننا من تقدير معالم الانحدار. افترض، مثلا، أن شركة تأمين ترغب في دراسة العلاقة بين إنتاجية محلليها في معالجة الدعاوى وبين مدة التدريب. وتستخدم الدراسة ثمانية محللين. نختار ثلاثة منهم بصورة عشوائية ليتمرنوا لمدة أسبوعين واثنين لثلاثة أسابيع وثلاثة لخمسة أسابيع. ثم نشاهد إنتاجيتهم خلال الأسابيع العشرة التالية. والبيانات التي نحصل عليها ستكون بيانات تجريبية لأن نوعا من السيطرة قد مُورس على المتغير المستقل وهو طول فترة التدريب.

وعندما تُمارس القيود على المتغير (المتغيرات) المستقل (المستقلة) من خلال تخصيص عشوائي، كما هو في مثال دراسة الإنتاجية، فإن البيانات التجريبية الناتجة تزودنا بمعلومات أقوى بكثير عن علاقات السبب والتأثير من تلك التي توفرها بيانات المشاهدة. والسبب في ذلك يعود إلى أن العشوائية تتولى موازنة تأثيرات المتغيرات الأخرى التي يمكن أن تؤثر في المتغير التابع، مثل تأثير استعدادات المستخدم على الإنتاجية.

وفي مصطلحات تصميم التجارب، تُسمّى فترة التدريب المخصصة للمحلل في مثال دراسة الإنتاجية *معالجة*. ويُسمى المحللون المشاركون في الدراسة *الوحدات التجريبية*. وعندئذٍ تتمثل السيطرة على المتغيرات المستقل بتخصيص معالجة لكل وحدة تجريبية بطرق عشوائية.

تصميم تام العشوائية

التصميم تام العشوائية هو النوع الأساسي في التصميم الإحصائي المتعلق بعملية تخصيص المعالجات عشوائيا للوحدات التجريبية (والعكس بالعكس). ووفق هذا التصميم، يتم التخصيص بالكامل عشوائيا. وتمنح هذه العشوائية التامة كل وحدة

تجريبية الفرصة نفسها في تلقّي أي من المعالجات، أو بصورة مكافئة، يكون لجميع الاختيارات الممكنة من الوحدات التجريبية التي خُصِّصت لمعالجات مختلفة، الفرصة نفسها.

وعلى وجه الخصوص، يكون التصميم تام العشوائية مفيدا عندما تكون الوحدات التجريبية متجانسة إلى حد كبير. وهذا التصميم مرن جدا، ويلائم أي عدد من المعالجات، ويسمح بأحجام عينات مختلفة لمعالجات مختلفة. وعيبه الرئيس هو أنه عندما تكون الوحدات التجريبية غير متجانسة، فلا يكون هذا التصميم فعالا بالمقارنة مع تصاميم إحصائية أخرى.

استخدام جدول الأرقام العشوائية أو مولد الأرقام العشوائية المنتظم للتعشية. تتطلب التعشية في تصميم تام العشوائية أن تأخذ الوحدات التجريبية (أو المعالجات) مواقعها وفق ترتيب عشوائي. ولتوضيح ذلك، لنعد ثانية إلى مثال دراسة الإنتاجية، فلدينا هنا ثلاث معالجات (T_1- تدريب لأسبوعين، T_2- تدريب لثلاثة أسابيع وT_3- تدريب لخمسة أسابيع) وتضم الدراسة ثمانية محلّلين بحيث يُخصص $n_1 = 3$ للمعالجة T_1، ويُخصص $n_2 = 2$ للمعالجة T_2، ويُخصص $n_3 = 3$ للمعالجة T_3. والحالة إذن هي كما يلي:

المعالجات T_1, T_2, T_3

حجوم العينات $n_1 = 3, n_2 = 2, n_3 = 3$

والآن نضع لائحة (بترتيب اختياري) بالمعالجات المراد تخصيصها للمحلّلين الثمانية:

$$T_1 \quad T_1 \quad T_1 \quad T_2 \quad T_2 \quad T_3 \quad T_3 \quad T_3$$

ولتعشية المعالجات على الوحدات التجريبية، نرقم المحللين من 1 إلى 8. وبالتالي نحتاج إلى الحصول على ترتيبات أو تباديل عشوائية للأرقام 1، 2،...،8. نقوم بذلك بالاستفادة إما من جدول للأرقام العشوائية ومن مولد الأرقام العشوائية المنتظم. في البداية نضع الأرقام الثمانية في تسلسلها الطبيعي:

1 2 3 4 5 6 7 8

وبعد ذلك نولد ثمانية أعداد عشوائية بين 000 و 999 (استخدمنا أعدادا ذات ثلاث مراتب لجعل احتمال تكرار العدد نفسه (صغيرا) ونكتبها فوق الأرقام 1 إلى 8، وحدث أن كانت الأرقام المولّدة كما يلي:

263	107	362	279	084	349	325	737
8	7	6	5	4	3	2	1

والآن نرتب أزواج الأرقام وفق القيم المتصاعدة للأعداد العشوائية:

737	362	349	325	279	263	107	084
1	6	3	2	5	8	7	4

وهكذا نحصل على التخصيص العشوائي التالي للمحللين الثمانية على المعالجات الثلاث:

المحلل:	4	7	8	5	2	3	6	1
المعالجة:	T_1	T_1	T_1	T_2	T_2	T_3	T_3	T_3

وبذلك يتلقى المحللون 4، 7 و 8 المعالجة T_1 (تدريب أسبوعين) وهكذا.

وإذا تكرر الحصول على العدد العشوائي نفسه، فباستطاعتنا اختيار أو توليد عدد عشوائي آخر كبديل عن العدد المكرر.

(٢ ـ ٥) نظرة عامة على تحليل الانحدار

يمكن الإفادة من تحليل الانحدار المعروض في هذا الفصل والفصول التالية في بيانات مشاهدة أو في بيانات تجريبية وفق تصميم تام العشوائية. (ونستطيع الانتفاع من تحليل الانحدار في بيانات من أنواع أخرى لتصميم التجارب ولكن نماذج الانحدار المعروضة هنا تحتاج عندئذٍ إلى تبديل.) وسواء أكان البيان بيان مشاهدات أو بيانا تجريبيا، فمن الضروري أن تكون شروط نموذج الانحدار مناسبة للبيانات التي في حوزتنا.

وسوف نبدأ دراستنا لتحليل الانحدار بالاستقراء عن معالم الانحدار في نموذج الانحدار الخطي البسيط (2.1). وفي الحالة النادرة التي تتوافر فيها معلومات سابقة أو نظرية تحدد لنا، بمفردها، نموذج الانحدار المناسب تكون الاستقراءات المبنية على نموذج الانحدار هذا هي الخطوة الأولى في تحليل الانحدار. وعلى كل حال، في الحالات

الاعتيادية حيث لا نملك المعلومات الكافية لتحديد نموذج الانحدار المناسب سلفا، تكون الدراسة الاستكشافية للبيانات الخطوة الأولى كما هو موضح في مخطط التدفق في الشكل (٢-٩). وبالاستناد على هذا التحليل الاستكشافي المبدئي يُستحدث نموذج أولي أو أكثر للانحدار. وتُفحص نماذج الانحدار هذه من حيث صلاحيتها للبيانات التي في حوزتنا ثم تُنقّح أو تُستحدث نماذج جديدة حتى يقتنع الدارس بأن نموذجا بالذات من بينها هو النموذج المناسب. وعندئذٍ فقط تتم الاستقراءات بالاستناد إلى نموذج الانحدار هذا، كالاستقراءات حول معالم الانحدار للنموذج أو تنبؤات بمشاهدات جديدة.

ولأسباب تربوية، نبدأ باستقراءات على أساس أن النموذج الانحداري اعتبر أخيرا أنه النموذج المناسب، وذلك قبل أن نتصدّى لكيفية تطوير نموذج انحدار مناسب، ولابد من فهم نماذج الانحدار وكيفية الاستفادة منها قبل أن يكون الدارس قادرا على فهم المسائل التي ينطوي عليها تطوير نموذج انحدار مناسب فهما تاما.

(٢-٦) تقدير دالة الانحدار

مثال.

سوف نستخدم مثال شركة وستوود لتوضيح تقدير دالة انحدار خطية بسيطة. ويقدم الجدول (٢-١) (ص ٤٦) بيانات مشاهدة لعشر دورات إنتاجية حديثة ويتضمن حجم دورة الانتاج (X) وعدد ساعات العمل المتطلبة للدورة. وسوف نرمز للمشاهدات (X, Y) في المحاولة الأولى بـ (X_1, Y_1) وفي المحاولة الثانية بـ (X_2, Y_2) وعموما في المحاولة بـ (X_i, Y_i) حيث $i = 1,2,...,n$ وفي بيانات الجدول (٢-١)، $X_1 = 30$، $Y_1 = 73$ وهكذا و $n = 10$.

طريقة المربعات الدنيا

لإيجاد مقدّرات "جيدة" لمعالم الانحدار β_0 و β_1 سوف نستخدم طريقة المربعات الدنيا. ولكل مشاهدة عينة (X_i, Y_i)، تأخذ طريقة المربعات الدنيا انحراف Y_i عن قيمته المتوقعة.

$$Y_i - (\beta_0 + \beta_1 X_i) \tag{2.7}$$

على وجه الخصوص، تتطلب طريقة المربعات الدنيا اعتبار مجموع مربعات الانحرافات الـ n ويرمز لهذا المعيار بـ Q :

$$Q=\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1 X_i)^2 \tag{2.8}$$

شكل (٢-٩) استراتيج نموذجي لتحليل الانحدار

البداية

↓

تحليل استكشافي للبيانات

↓

طور نموذج انحدار تجريبي أو أكثر

↓

هل يتلاءم واحد أو أكثر من نماذج الانحدار مع ما لدينا من بيانات

لا ← قم بتعديل نماذج الانحدار و / أو طوّر نماذج جديدة

نعم ↓

حدد النموذج الأفضل

↓

قم باستقراءات معتمدا على نموذج الانحدار

↓

قف

وطبقا لطرق المربعات الدنيا، تكون تقديرات β_0 و β_1 تلك القيم b_0 و b_1، على الترتيب، التي تجعل المعيار أصغر ما يمكن، وذلك من أجل مشاهدات العينة $(X_i\ ,Y_i)$ المعطاة.

جدول (٢-١) بيانات حجم الدفعة وعدد ساعات العمل ـ مثال شركة وستوود

دورة الإنتاج	حجم الدفعة	ساعات العمل
1	30	73
2	20	50
3	60	128
4	80	170
5	40	87
6	50	108
7	60	135
8	30	69
9	70	148
10	60	132

مثال. يعيد الشكل (٢-١٠)ا رسم انتشار بيانات العينة من الجدول (٢-١) لمثال شركة وستوود. وفي الشكل (٢-١٠)ب رسم توفيقي لخط انحدار استُخدمت فيه التقديرات الاختيارية.

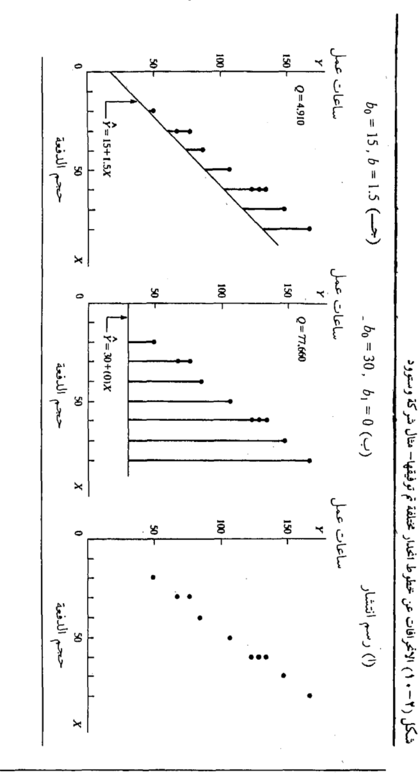
$$b_1 = 0 \quad , \quad b_0 = 30$$

ويبين الشكل (٢-١٠)ب كذلك الانحرافات $Y_i - 30 - (0)X_i$ لاحظ أن كل انحراف يقابل المسافة العمودية بين Y_i والخط التوفيقي للانحدار. وواضح أن التوافق ضعيف ومن ثم فإن الانحرافات كبيرة وكذلك مربعات الانحرافات. ومجموع مربعات الانحرافات (المشاهدات مرتبة تصاعديا) هو:

$$Q = (50 - 30)^2 + (69 - 30)^2 + ... + (170 - 30)^2 = 77660$$

ويحوي الشكل (٢-١٠)جـ الانحرافات $Y_i - b_0 - b_1 X_i$ من أجل التقديرات $b_0 = 15$ و $b_1 = 1.5$. التوافق، هنا، أحسن (ومع ذلك مازال دون الجيّد) والانحرافات أصغر بكثير، وبالتالي هبط مجموع مربعات الانحرافات إلى $Q = 4910$. وهكذا يقابل حسن مطابقة خط الانحدار للبيانات مجموع صغير لِـ Q.

والهدف من طريقة المربعات الدنيا، هو إيجاد تقديرات b_0 و b_1 لِـ β_0 و β_1 على الترتيب، يكون Q من أجلها أصغر مايمكن. وبمعنى، ستأتي مناقشته بعد قليل، تزوّدنا هذه التقديرات بتوفيق "جيد" لدالة انحدار خطي.



شكل (٢-١٠) الانحرافات عن خطوط انحدار مختلفة تم توفيقها- مثال شركة وستوود

مقدرات المربعات الدنيا. يمكن الحصول على المقدّرات b_0 و b_1، التي تحقق قاعدة المربعات الدنيا، بطريقتين أساسيتين. في الأولى يمكن استخدام طرق البحث العددية التي تحسب بصورة متناسقة معيار المربعات الدنيا Q لتقديرات b_0 و b_1 مختلفة حتى نعثر على تلك التي تجعل Q أصغر مايمكن. وفي الطريقة الثانية نجد بأسلوب تحليلي القيم b_0 و b_1 التي تجعل Q أصغر مايمكن. وتكون الطريقة التحليلية ممكنة عندما لايكون النموذج معقدا من الناحية الرياضية، كما هو الحال هنا.

ويمكن إثبات أن القيم b_0 و b_1 التي تجعل Q أصغريا، لأي مجموعة محددة من بيانات عينة، تُعطى بالمعادلتين الآنيتين التاليتين:

$$\sum Y_i = nb_0 + b_1 \sum X_i \tag{2.9a}$$

$$\sum X_i Y_i = b_0 \sum X_i + b_1 \sum X_i^2 \tag{2.9b}$$

وتُسمى المعادلتان (2.9a) و(2.9b) المعادلتين الناظميتين، وتسمى b_0 و b_1 *المقدرات النقطية* لـ β_0 و β_1 على الترتيب.

وتُحسب الكميات $\sum X_i$ و $\sum Y_i$ وشبيهاتها في (2.9) من مشاهدات العينة (X_i, Y_i). ومن ثَمَّ يمكن حل المعادلتين معا من أجل b_0 و b_1 وكبديل آخر يمكن الحصول على b_0 و b_1 مباشرة كما يلي:

$$b_1 = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \tag{2.10a}$$

$$b_0 = \frac{1}{2}\left(\sum Y_i - b_1 \sum X_i\right) = \bar{Y} - b_1 \bar{X} \tag{2.10b}$$

حيث $\bar{X}$ و $\bar{Y}$ متوسطا المشاهدات X_i و Y_i على الترتيب.

ملاحظة

يمكن اشتقاق المعادلتين الناظميتين (2.9) باستخدام حساب التفاضل والتكامل. فمن أجل مشاهدات (X_i, Y_i) معطاة تكون الكمية Q في (2.8) دالة في B_0 و B_1.

ويمكن الحصول على قيم B_0 و B_1 التي تجعل Q أصغر ما يمكــن باشـتقاق (2.8) جزئيـا بالنسبة لِـ β_0 و β_1، لنجد:

$$\frac{\partial Q}{\partial \beta_0} = -2\sum(Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum X_i(Y_i - \beta_0 - \beta_1 X_i)$$

ثم نضع المشتقتين الجزئيتين مســاويتين للصفـر، مسـتخدمين b_0 و b_1 كرمزيـن لقيم β_0 و β_1، على الترتيب، اللتين تجعلان Q أصغر ما يمكن لنحصل على:

$$-2\sum(Y_i - b_0 - b_1X_i) = 0$$

$$-2\sum X_i(Y_i - b_0 - b_1X_i) = 0$$

ونحصل بالتبسيط على:

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i) = 0$$

$$\sum_{i=1}^{n}X_i(Y_i - b_0 - b_1X_i) = 0$$

وبفك القوسين لدينا:

$$\sum Y_i - nb_0 - b_1\sum X_i = 0$$

$$\sum X_iY_i - b_0\sum X_i - b_1\sum X_i^2 = 0$$

ومنها نحصل على المعادلتين الناظميتين (2.9) وذلك بإعادة ترتيب الحدود.

وباختبار المشتقات الجزئية الثانية نرى أن القيمة الصغرى تتحقق عنـد مقـدِّري المربعات الدنيا b_0 و b_1.

خـواص مقـدرات المربعـات الدنيـا. تعـرض نظريـة مهمـة تسـمى نظريـة جـــاوس – ماركوف مايلي:

(2.11) تحت شروط نموذج الانحدار (2.1) تكون مقدَّرات المربعـات الدنيـا b_o و b_1 في (2.10) غير منحازة ولها أصغر تباين بين كافة المقدرات الخطية غـير المنحـازة.

وتصرح هذه النظرية، والتي سوف تُبرهن في الفصل القادم، أولا، أن كـلا مـن b_0 و b_1 مقدِّران غير منحازين. وهكذا نكتب:

$$E\{b_0\} = \beta_0$$

$$E\{b_1\} = \beta_1$$

ولذلك لايميل المقدّر، ميلا منتظما، إلى التقدير بالزيادة أو التقدير بالنقصان.

ثانيا، تعرض النظرية أن تباين توزيع المعاينة لِ b_0 و b_1 أقل من تبـاين أي مـن المقدرات الأخرى التي تنتمي إلى صف خاص من المقدرات. وهكذا، تكـون مقـدّرات المربعات الدنيا أكثر إحكاما من أي من هذه المقدّرات. ويتألف صف المقدّرات، الذي تتصدره مقدرات المربعات الدنيا كأفضل مافيه، من جميع المقدرات غير المنحـازة والـتي هـي دوال خطيـة في المشـاهدات Y_1،...،Y_n والمقـدران b_0 و b_1 همـا دالتـان مـن هـذا الصف أي دالتان خطيتان في الـY_i اعتبر، مثلا،b_1 فلدينا من (2.10a).

$$b_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}$$

وسنبيّن في الفصل الثالث أن هذه العبارة تساوي إلى:

$$b_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} = \sum k_i Y_i$$

حيث:

$$k_i = \frac{(X_i - \overline{X})}{\sum(X_i - \overline{X})^2}$$

وحيث إن الـ k_i ثوابت معروفة (لأن الـ X_i ثوابت معروفة)، يكون b_1 تركيبا خطيا في الـ Y_i، وبالتالي فهو مقدَّر خطي. وبالطريقة نفسها يمكن إثبات أن b_0 مقـدر خطي. ومن بين جميع المقدرات الخطية غير المنحـازة، فـإن b_0 و b_1 همـا الأقـل تباينـا عند سحب عينات متكررة يبقى مستوى X فيها ثابتا.

مثال. لتوضيح حساب مقدرات المربعات الدنيا b_0 و b_1 سوف نستخدم مثـال شركة وستوود. يقدم الجدول (٢-١) بيانات العينة وهي مرسـومة في الشـكل (٢-١٠)أ؛ ويعطي الجدول (٢-٢) النتائج الأساسية المطلوبة لحساب b_0 و b_1. ولدينا:

$\sum X_i^2 = 28400$ ، $\sum X_i Y_i = 61800$ ، $\sum X_i = 500$ ، $\sum Y_i = 1100$

و $n = 10$ ونحصل من استخدام (2.10) على:

$$b_1 = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{61800 - \frac{500(1100)}{10}}{28400 - \frac{(500)^2}{10}} = 2.0$$

$$b_0 = \frac{1}{n}\left(\sum Y_i - b_1 \sum X_i\right) = \frac{1}{10}[1100 - 2.0(500)] = 10.0$$

وهكذا قدّرنا أن متوسط عدد ساعات العمل يزداد بـ 2.0 ساعة لكل وحدة يزدادها حجم الدفعة.

تقدير نقطي لمتوسط الاستجابة

دالة الانحدار المقدّرة. إذا عُرفت مقدرات العينة b_0 و b_1 للمعلمتين في دالة الانحدار (2.3):

$$E\{Y\} = \beta_0 + \beta_1 X$$

فإننا نقدر دالة الانحدار كما يلي:

$$\hat{Y} = b_0 + b_1 X \tag{2.12}$$

حيث $\hat{Y}$ (تُقرأ Y قبّعة) هي القيمة المقدرة لدالة الانحدار عند المستوى X للمتغير المستقل.

سوف نسمي *قيمة* متغير الإستجابة، "*استجابة*"، ونسمي $E\{Y\}$ *متوسط الاستجابة*. وهكذا يكون متوسط الاستجابة هو متوسط التوزيع الاحتمالي لـ Y المقابل للمستوى X للمتغير المستقل، وعندئذ يكون $\hat{Y}$ مقدّرا نقطيا لمتوسط الاستجابة عندما يكون مستوى المتغير المستقل X. وكامتدادٍ لنظرية جاوس ـ ماركوف (2.11) يمكن إثبات أن $\hat{Y}$ مقدر غير منحاز لـ $E\{Y\}$ بتباين أصغري في صف المقدرات الخطية وغير المنحازة لـ $E\{Y\}$.

وفي الحالات قيد الدراسة، سوف نسمي $\hat{Y}_i$ حيث:

$$\hat{Y}_i = b_0 + b_1 X_i \qquad i = 1,\dots,n \tag{2.13}$$

القيمة التوفيقية للمشاهدة i. وهكذا ننظر الى القيمة التوفيقية $\hat{Y}_i$ كشيء مميز عن القيمة الملحوظة.

مثال. في مثال شركة وستوود، وجدنا أن مقدرات المربعات الدنيا لمعاملات الانحدار كانت:

$$b_0 = 10.0 \qquad b_1 = 2.0$$

ومن ثَمَّ فإن دالة الانحدار المقدَّرة هي:

$$\hat{Y} = 10.0 + 2.0X$$

جدول (٢-٢) الحسابات الأساسية للحصول على b_0 و b_1 - مثال شركة وستوود

i	Y_i	X_i	X_iY_i	X_i^2	Y_i^2
1	73	30	2190	900	5329
2	50	20	1000	400	2500
3	128	60	7680	3600	16384
4	170	80	13600	6400	28900
5	87	40	3480	1600	7569
6	108	50	5400	2500	11664
7	135	60	8100	3600	18225
8	69	30	2070	900	4761
9	148	70	10360	4900	21904
10	132	60	7920	3600	17424
المجموع	1100	500	61800	28400	134660

وإذا كنا نهتم بمتوسط عدد ساعات العمل عندما يكون حجم الدفعة $X=55$ فإن تقديرنا النقطي هو:

$$\hat{Y} = 10.0 + 2.0(55) = 120$$

وهكذا، نقدر أن متوسط عدد ساعات العمل لدورات إنتاج من $X=55$ وحدة هو 120. وتفسير ذلك يعني أنه إذا أُنتجت دورات كثيرة حجمها 55، تحت شروط الدورات العشر في العينة، فإن متوسط زمن العمل لهذه الدورات العديدة يقارب 120 ساعة. وبالطبع، فإنه من المرجح أن يقع وقت العمل لدورة واحدة من الحجم 55 فوق أو تحت متوسط الاستجابة، وذلك بسبب خاصية التغير المتأصلة في النظام والممثلة بحد الخطأ في النموذج.

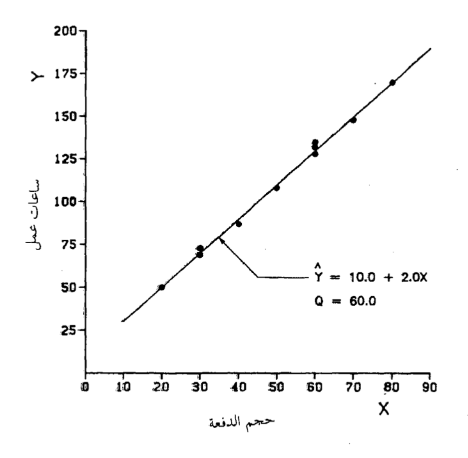
ويحوي الشكل (٢-١١) رسم حاسب آلي لدالة الانحدار المقدَّرة $\hat{Y} = 10.0 + 2.0X$ بالإضافة إلى البيانات الأصلية. لاحظ التحسن في توفيق خط انحدار المربعات الدنيا مقارنة بالخطوط الكيفية في الشكل (٢-١٠). وفي الواقع سنثبت، بعد قليل، أن المعيار Q في الانحدار الخطي للمربعات الدنيا ماهو الآن إلا $Q = 60$ وهي قيمة أصغر بكثير من قيم Q لخطوط التوفيق الكيفية في الشكل (٢-١٠).

ونحصل على القيم التوفيقية لبيانات العينة بتعويض قيم X في العينة في دالة الانحدار المقدَّرة.. فمثلا في عينتنا، $X_1 = 30$. وبالتالي تكون القيمة التوفيقية للمشاهدة الأولى:

$$\hat{Y}_1 = 10.0 + 2.0(30) = 70$$

وتُقارن هذه مع ساعات العمـل الملحوظـة 73 = Y_1. يحـوي الجـدول (٢-٣) (ص٥٢) جميـع القيـم الملحوظـة والقيـم التوفيقيـة لبيانـات شـركة وسـتوود وذلـك في العمودين 2 و 3 على الترتيب.

شكل (٢-١١) المشاهدات وخط انحدار المربعات الدنيا لمثال شركة وستوود: $b_0 = 10.0$ و $b_1 = 2.0$



النموذج (2.6) كنموذج بديل. اذا استخدم النموذج البديل (2.6):

$$Y_i = \beta_0^* + \beta_1(X_i - \bar{X}) + \varepsilon_i$$

يبقى مقدر المربعات الدنيا b_1 لـ β_1 كما كان سابقا وباستخدام (2.10b) يكون مقدر المربعات الدنيا لـ $\beta_0^* = \beta_o + \beta_1 \overline{X}$:

$$b_0^* = b_0 + b_1\overline{X} = (\overline{Y} - b_1\overline{X}) + b_1\overline{X} = \overline{Y} \qquad (2.14)$$

وبالتالي فإن معادلة الانحدار المقدرة للنموذج البديل (2.6):

$$\hat{Y} = \overline{Y} + b_1(X - \overline{X}) \qquad (2.15)$$

وفي مثال شركة وستوود $\overline{Y} = 1100 / 10 = 110$ و $\overline{X} = 500/10 = 50$ [جدول (٢-٢)] وبالتالي يصبح الشكل البديل لدالة الانحدار المقدرة:

$$\hat{Y} = 110.0 + 2.0(X - 50)$$

ومن أجل المشاهدة الأولى في عينتا $X_1 = 30$، نقدر متوسط الاستجابة بـ:

$$\hat{Y}_1 = 110.0 + 2.0(30\text{-}50) = 70$$

وهو بالطبع متطابق مع نتيجتنا السابقة.

جدول (٢-٣) القيم التوفيقية والرواسب ومربعات الرواسب – مثال شركة وستوود

دورة إنتاج i	(١) حجم دفعة X_i	(٢) ساعات عمل Y_i	(٣) متوسط استجابة مقدر	(٤) راسب $Y_i - \hat{Y}_i = e_i$	(٥) مربع راسب $(Y_i - \hat{Y}_i) = e_i^2$
1	30	73	70	3+	9
2	20	50	50	0	0
3	60	128	130	2-	4
4	80	170	170	0	0
5	40	87	90	3+	9
6	50	108	110	2-	4
7	60	135	130	5+	25
8	30	69	70	1-	1
9	70	148	150	2-	4
10	60	132	130	2+	4
المجموع	500	1,100	1,100	0	60

الرواسب

الراسب الـ i هو الفرق بين القيمة الملحوظة Y_i والقيمة التوفيقية المقابلة $\hat{Y}_i$.

وإذا رمزنا بـ e_i لهذا الراسب يمكننا كتابة:

$$e_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i \tag{2.16}$$

ويوضح الشكل (٢-١٢) الرواسب العشرة في مثال شركة وستوود. وقد أوضحت حجوم الرواسب بخطوط عمودية تصل بين قيمة Y الملحوظة والقيمة التوفيقية على خط الانحدار المقدّر وقد حُسبت الرواسب في العمود 4 من الجدول (٢-٣).

وينبغي لنا التمييز بين قيمة حد الخطأ في النموذج $\varepsilon_i = Y_i - E\{Y_i\}$ والراسب $e_i = Y_i - \hat{Y}_i$. فالأول يعني انحراف Y_i الرأسي عن خط الانحدار الحقيقي غير المعروف، وبالتالي فهو غير معروف. ومن جهة أخرى فإن الراسب هو الانحراف الرأسي لـ Y_i عن القيمة التوفيقية $\hat{Y}_i$، على خط الانحدار المقدّر.

والرواسب مفيدة جدا في دراسة ماإذا كان نموذج انحدار معيّن مناسبا للبيانات التي في حوزتنا. وسوف نناقش مثل هذا الاستخدام في الفصل الرابع.

خواص خط الانحدار التوفيقي

لخط الانحدار الموفّق بطرق المربعات الدنيا عدد من الخواص التي تستحق الذكر:

١- مجموع الرواسب يساوي صفرا:

$$\sum_{i=1}^{n} e_i = 0 \tag{2.17}$$

ويوضح العمود الرابع في الجدول (٢-٣) هذه الخاصية لمثال شركة وستوود. وبالطبع قد تحدث أخطاء نتيجة تدوير الأرقام العشرية لأي مشاهدة معينة مما يجعل مجموع الرواسب غير مساوٍ للصفر تماما.

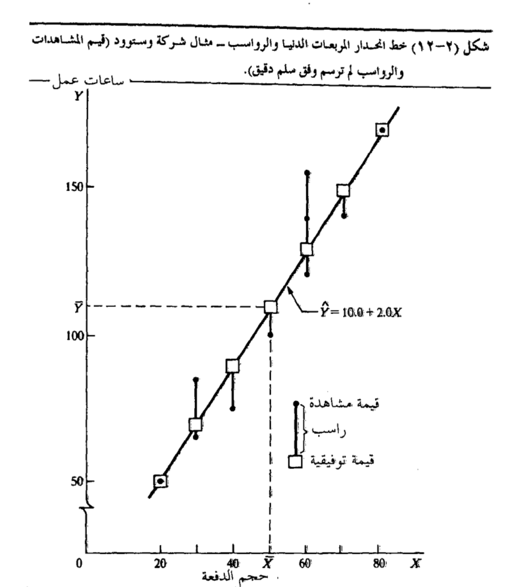
٢- يكون مجموع مربعات الرواسب $\sum e_i^2$ أصغريا. وهذا هو المتطلب الذي كان ينبغي تحققه عند استنباط مقدّرات المربعات الدنيا لمعالم الانحدار.

٣- مجموع القيم الملحوظة Y_i يساوي مجموع القيم التوفيقية $\hat{Y}_i$:

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i \tag{2.18}$$

وقد وضّحت هذه الخاصية في العمودين ٢ و ٣ من الجدول (٢-٣) لمثال شركة وستوود. وهذا يستتبع كون متوسط الـ $\hat{Y}_i$ هو ذاته متوسط الـY_i وبالتحديد.

شكل (٢-١٢) خط انحدار المربعات الدنيا والرواسب ــ مثال شركة وستوود (قيم المشاهدات والرواسب لم ترسم وفق سلم دقيق).



٤- يكون مجموع الرواسب الموزونة صفرا عندما يوزن راسب المشاهدة i بمستوى المتغير المستقل في تلك المشاهدة:

$$\sum_{i=1}^{n} X_i e_i = 0 \qquad (2.19)$$

٥- يكون مجموع الرواسب المرجحة صفرا عندما يوزن راسب المشاهدة i بالقيمة التوفيقية لمتغير الاستجابة للمشاهدة i:

$$\sum_{i=1}^{n} \hat{Y}_i e_i = 0 \qquad (2.20)$$

٦- يمر خط الانحدار دائما بالنقطة $(\overline{X}, \overline{Y})$. ويوضح الشكل (٢-١٢) هذه الخاصّية لمثال شركة وستوود.

تعليقات

١- يمكن استنتاج الخواص الست للرواسب مباشرة من المعادلات الناظمية (2.9) للمربعات الدنيا. فمثلا، تبرهن الخاصية في (2.17) كما يلي:

$$\sum e_i = \sum (Y_i - b_0 - b_1 X_i) = \sum Y_i - nb_0 - b_1 \sum X_i = 0$$

بالاستفادة من المعادلة الناظمية الأولى (2.9a).

ويمكن بسهولة إثبات الخاصية ٦، وهي أن خط الانحدار يمر دائما بالنقطة $(\overline{X}, \overline{Y})$، وذلك من الشكل البديل (2.15) لخط الانحدار المقدَّر. فعندما يكون $X = \overline{X}$ نجد:

$$\hat{Y} = \overline{Y} + b_1 (X - \overline{X}) = \overline{Y} + b_1 (\overline{X} - \overline{X}) = \overline{Y}$$

٢- تنطبق خواص الرواسب التي لاحظناها هنا على نموذج الانحدار (2.1). ولاتنطبق هذه الخواص على جميع نماذج الانحدار، كما سنلاحظ في الفصل الخامس.

(٢-٧) تقدير تباين حدود الأخطاء σ^2

ولأغراض متعددة نحتاج إلى تقدير التباين σ^2 لحدود الأخطاء ε_i في نموذج الانحدار (2.1) فكثيرا ما نرغب في الحصول على مؤشر عن تباين التوزيع الاحتمالي لـ Y. وبالإضافة إلى ذلك وكما سنرى في الفصل القادم، يتطلب العديد من الاستقراءات حول دالة الانحدار والتنبؤ عن Y، تقدير σ^2.

تقدير نقطي لِـ σ^2

مجتمع بمفرده. كي نضع الأساس لتطوير مقدِّر لِـ σ^2 في نموذج الانحدار (2.1)، دعنا نعتبر لهنيهة المسألة الأبسط وهي المعاينة من مجتمع بمفرده. وللحصول على تباين العينة s^2 نبدأ باعتبار انحراف المشاهدة Y_i عن المتوسط المقدر $\overline{Y}$ فنربعه، ثم نجمع جميع هذه الانحرافات المربعة:

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

ويُسمى مجموع كهذا *مجموع مربعات*. وبعد ذلك نقسم مجموع المربعات على درجات الحرية المرتبطة به. والعدد هنا هو $(n-1)$ فقد خسرنا درجة حرية واحدة نظرا لاستخدام المقدّر $\bar{Y}$ بدلا من متوسط المجتمع. ويكون تباين العينة المعتاد هو المقدّر الناتج:

$$s^2=\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{n-1}$$

وهو مقدّر غير منحاز للتباين σ^2 لمجتمع لانهائي. وغالبا مايسمى تباين العينة متوسط مربعات، لأن مجموع المربعات قُسم على عدد درجات الحرية المناسبة.

نموذج الانحدار. لا يختلف المنطق في تطويرمقدّر لِ σ^2 في نموذج الانحدار عنه عند المعاينة من مجتمع منفرد. ولنتذكر، لهذا الغرض، من (2.4) أن تباين كل مشاهدة Y_i هو σ^2 وهو نفسه لكل حد خطأ ε_i. ونحتاج مرة ثانية حساب مجموع مربعات الانحرافات، ولكن ينبغي إدراك أن الـ Y_i، جاءت من توزيعات احتمالية مختلفة لها متوسطات تختلف باختلاف المستوى X_i. وهكذا ينبغي حساب انحراف مشاهدة Y_i عن متوسطها المقدَّر الخاص بها $\hat{Y}_i$. ومن ثم تكون الانحرافات هي الرواسب:

$$Y_i-\hat{Y}_i=e_i$$

ويكون مجموع المربعات المناسب ويرمز له بـ SSE:

$$SSE=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2=\sum_{i=1}^{n}(Y_i-b_0-b_1X_i)^2=\sum_{i=1}^{n}e_i^2 \qquad (2.21)$$

حيث يرمز SSE لمجوع مربعات الخطأ أو لمجموع مربعات الرواسب.

ويرتبط بمجموع المربعات SSE، $n-2$ من درجات الحرية. وخسرنا درجتي حرية لأنه كان علينا تقدير كل من β_0 و β_1 من أجل الحصول على المتوسطات المقدَّرة $\hat{Y}_i$. وبالتالي فإن متوسط المربعات المناسب ويُرمز له بـ MSE هو:

$$MSE=\frac{SEE}{n-2}=\frac{\sum(Y_i-\hat{Y}_i)^2}{n-2}=\frac{\sum(Y_i-b_0-b_1X_i)^2}{n-2}=\frac{\sum e_i^2}{n-2} \qquad (2.22)$$

حيث يرمز MSE لمتوسط مربعات الخطأ أو متوسط مربعات الرواسب.

ويمكن إثبات أن MSE مقدر غير منحاز لِ σ^2 في نموذج الانحدار (2.1):

$$E\{MSE\} = \sigma^2 \tag{2.23}$$

ويكون مقدر الانحراف المعياري ببساطة الجذر التربيعي الموجب لـ MSE.

صيغ حسابية بديلة.

يوجدعدد من الصيغ الحسابية البديلة لـ SSE. وفيما يلي ثلاث صيغ منها:

$$SSE = \sum Y_i^2 - b_0 \sum Y_i - b_1 \sum X_i Y_i \tag{2.24a}$$

$$SSE = \sum (Y_i - \overline{Y})^2 - \frac{\left[\sum (X_i - \overline{X})(Y_i - \overline{Y})\right]^2}{\sum (X_i - \overline{X})^2} \tag{2.24b}$$

$$SSE = \left[\sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}\right] - \frac{\left(\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right)^2}{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}} \tag{2.24c}$$

تعليقات

١– تكون الصيغة (2.24a) مفيدة إذا تم حساب b_0 و b_1 وخلاف ذلك تكون (2.24b) و (2.24c) الأكثر مباشرة.

٢– في (2.24a) يجب حساب المقدرات b_0 و b_1 إلى عدد كبير من الأرقام العشرية كي نحصل على نتائج موثوقة لـ SSE.

٣– لاتوفر أي من الصيغ البديلة الثلاث الرواسب e_i بصورة صريحة. وكما ذكرنا سابقا، فإن الرواسب مفيدة في دراسة صلاحية النموذج أو مصداقيته.

مثال

وبالعودة إلى مثال شركة وستوود، سنحسب SSE مستخدمين (2.21). وقد حصلنا على الرواسب سابقا في العمود (٤) من الجدول (٢–٣). ويبين هذا الجدول أيضا مربع الرواسب في العمود (٥) ونحصل من هذه النتائج على:

$$SSE = 60$$

وبما أن عدد درجات الحرية المرتبطة بـ SSE هو $10 - 2 = 8$، فنجد:

$$MSE = \frac{60}{8} = 7.5$$

وأخيرا يكون $\sqrt{7.5} = 2.74$ ساعة عمل مقدرا نقطيا لـ σ، الانحراف المعيـاري للتوزيـع الاحتمالي لِـ Y، وذلك أيا كانت قيمة X .

لنعتبر مرة أخرى الحالة التي يكون حجم الدفعة فيها $X = 55$ وحدة. فقد سبق وأن قدرنا أن متوسط التوزيع الاحتمالي لِـ Y، المقابل لحجـم الدفعـة هـذا، هـو 120 ساعة عمل، والآن لدينا المعلومة الإضافية أن تقدير الانحراف المعياري لهذا التوزيع هـو 2.74 ساعة عمل.

إذا رغبنا باستخدام (2.24a)، مثلا، لحساب SSE، فإننا نحتاج إلى $\sum Y_i^2$ وهذا المجموع محسوب في الجدول (٢-٢). ومن ثم نحصل، باستخدام نتائج الجـدول (٢-٢) والتقديرات $b_0 = 10.0$ و $b_1 = 2.0$ على:

$$SSE = \sum Y_i^2 - b_0 \sum Y_i - b_1 \sum X_i Y_i$$
$$= 134660 - 10.0(1100) - 2.0(61800) = 60$$

وهي بالطبع النتيجة نفسها التي حصلنا عليها سابقا (باسـتثناء مـاقد ينتـج عـن تدويـر الأرقام العشرية).

(٢-٨) نموذج انحدار بخطأ طبيعي

مهما يكن الشكل الدالي لتوزيع ε_i (ومن ثَمَّ لـِ Y_i)، توفـر طريقـة المربعـات الدنيا مقدرات نقطية غير منحازة لِـ β_0 و β_1 لها تباين أصغـري بـين جميـع المقـدرات الخطية غير المنحازة. ولصياغة تقديرات بفترة أو القيام باختبارات سنحتاج، علـى أي حال، لوضع افتراض حـول الشـكل الـدالي لتوزيـع الـ ε_i. والافـتراض المعتـاد هـو أن حدود الخطأ موزعة طبيعيا، وسنتبنى هنا هذا الافتراض. ويسهّل حـد الخطـأ الطبيعـي نظرية تحليل الانحدار تسهيلا كبيرا وله مايبرره في العديد مـن الحـالات في واقـع الحيـاة التي يُطبّق فيها تحليل الانحدار.

النموذج

يُعرّف نموذج الانحدار الطبيعي كما يلي:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (2.25)$$

حيث:

Y_i هي الاستجابة الملحوظة في المشاهدة i.

X_i ثابت معروف، مستوى المتغير المستقل في المحاولة i.

β_0 و β_1 معلمتان

ε_i مستقلة و $N(0, \sigma^2)$ ، $i = 1, 2, \dots, n$

تعليقات

١- يمثل الرمز $N(0, \sigma^2)$ "موزعة طبيعيا"، بمتوسط 0 وتباين σ^2.

٢- نموذج الخطأ الطبيعي (2.25) هو ذاته نموذج الانحدار (2.1) بتوزيع خطأ غير محدد، ماعدا أن النموذج (2.25) يفترض أن الأخطاء ε_i موزعة طبيعيا.

٣- حيث يفترض نموذج الانحدار (2.25) أن الأخطاء موزعة طبيعيا فإن فرضية عدم الارتباط للأخطاء ε_i في نموذج الانحدار (2.1) تصبح فرضية استقلال في نموذج الخطأ الطبيعي.

٤- يتضمن نموذج الانحدار (2.25) أن الـY_i متغيرات عشوائية مستقلة وطبيعية، بمتوسط $E\{Y_i\} = \beta_0 + \beta_1 X_i$ وتباينσ^2، ويصور الشكل (٢-٤) نموذج الخطأ الطبيعي هذا، حيث يتوزع كل توزيع احتمالي لـِ Y طبيعيا بتباين ثابت ودالة انحدار خطية.

٥- والسبب الرئيس لتبرير فرضية طبيعية حدود الخطأ في العديد من الحالات هو أن حدود الخطأ تمثل أكثر ماتمثل، تأثيرات لعديد من العوامل التي لايذكرها النموذج صراحة وهي تؤثر إلى حد ما في متغير الاستجابة، ولايتوقف تغيرها العشوائي على المتغير المستقل X. وعلى سبيل المثال نجد في مثال شركة وستوود أن تأثير عوامل مثل الزمن المنصرم منذ دورة الإنتاج السابقة، والآلة المستخدمة بالذات، والفصل من السنة والعمال المستخدمين، ربما تغيرت كثيرا أو قليلا من دورة إلى دورة تغيرا عشوائيا ومستقلا عن حجم الدفعة. وكذلك، قد تكون هناك أخطاء قياس عشوائية في تسجيل Y. ومما تقدم وحيث إن لهذه التأثيرات العشوائية درجة من الاستقلال فيما بينها، فإن حد الخطأ المركب ε_i الذي يمثل

جميع هذه العوامل يميل للإذعان لنظرية النهاية المركزية فيتقارب توزيع حد الخطأ إلى الطبيعي عندما يصبح عدد العوامل المؤثرة كبيرا.

وسبب آخر للتبرير المتواتر لافتراض طبيعية حدود الخطأ، هو استناد طرق التقدير والاختبارات التي سنناقشها في الفصل التالي على توزيع t، وهو توزيع غير حساس لحيدان معتدل عن الطبيعية. وهكذا فإنه إذا لم يكن الحيدان عن الطبيعية خطيرا، وخصوصا فيما يتعلق بالالتواء، فإن معامل الثقة الحقيقي ومخاطر التورط بأخطاء ستكون قريبة من المستويات الموافقة لتوزيع طبيعي بالضبط.

تقدير المعالم بطرق الإمكانية العظمى

عندما يتحدد الشكل الدالي للتوزيع الاحتمالي لحدود الأخطاء، يمكن الحصول على مقدرات للمعالم β_0، β_1 وσ^2 بطريقة الإمكانية العظمى. وتستخدم هذه الطريقة التوزيع الاحتمالي المشترك لمشاهدات العينة. وعندما يُنظر إلى هذا التوزيع الاحتمالي المشترك كدالة في المعالم، علما أن مشاهدات العيّنة معروفة، فتدعى عندئذ دالة الإمكانية. وتكون دالة الإمكانية من أجل الانحدار بأخطاء طبيعية، علما أن مشاهدات العينة هي Y_1،...، Y_n كما يلي:

$$L(\beta_0,\beta_1,\sigma^2)=\prod_{i=1}^{n}\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left[-\frac{1}{2\sigma^2}(Y_i-\beta_0-\beta_1X_i)^2\right]$$

$$=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)^2\right] \quad (2.26)$$

وقيم β_0 ، β_1 و σ^2 التي تجعل دالة الإمكانية هذه أعظم ما يمكن هي مقدّرات الإمكانية العظمى، وهي:

(2.27)

المعلمة	مقدر الإمكانية العظمى
β_0	b_0 مثل $(2.10b)$
β_1	b_1 مثل $(2.10a)$
σ^2	$\hat{\sigma}^2=\frac{\sum(Y_i-\hat{Y}_i)^2}{n}$

وهكذا تكون مقدرات الإمكانية العظمى لِ β_0 و β_1 هي المقدرات نفسها التي توفّرها طريقة المربعات الدنيا. و مقدر الإمكانية العظمى $\hat{\sigma}^2$ منحاز، و من المعتاد استخدام المقدر MSE غير المنحاز كما هو معطى بـ (2.22). و نلاحظ أن المقدر غير المنحاز MSE يختلف اختلافا طفيفا عن مقدر الإمكانية العظمى $\hat{\sigma}^2$، خاصة عندما لا يكون n صغيرا:

$$MSE = \frac{n}{n-2}\hat{\sigma}^2 \tag{2.28}$$

تعليقات

١- حيث إن مقدرات الإمكانية العظمى b_0 و b_1 هي بالذات مقدرات المربعات الدنيا، فإن لها جميع خواص مقدرات المربعات الدنيا:

(أ) - إنها غير منحازة.

(ب) - لها أقل تباين بين جميع المقدرات الخطية غير المنحازة.

و بالإضافة إلى ذلك تمتلك مقدرات الإمكانية العظمى b_0 و b_1 لنموذج الانحدار بأخطاء طبيعية (2.25) خواص أخرى مرغوبة هي:

(جـ) - إنها متسقة، كما هو معرّف في (1.49).

(د) - إنها كافية، كما هو معرّف في (1.50).

(هـ) - إنها غير منحازة أصغرية التباين؛ أي لها أقل تباين ضمن صف جميع المقدرات غير المنحازة (خطية و غيرها).

و هكذا من أجل نماذج الخطأ الطبيعي، يمتلك المقدّران b_0 و b_1 العديد من الخواص المرغوبة.

٢- نحصل على القيم β_0 ، β_1 و σ^2 التي تجعل دالة الإمكانية L في (2.26) أعظم ما يمكن بأخذ المشتقات الجزئية لـ L بالنسبة لـ β_0 ، β_1 و σ^2 ومساواة كل منها بالصفر ثم حل نظام المعادلات الناتج. و يمكن استخدام $log_e L$، بدلا من L، لأن كلا من L و $log_e L$ تبلغ قيمتها العظمى عند القيم نفسها لِ β_0، β_1 و σ^2.

$$log_e L = -\frac{n}{2} log_e 2\pi - \frac{n}{2} log_e \sigma^2 - \frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2 \qquad (2.29)$$

وتكون المشتقات الجزئية لهذه الإمكانية اللوغاريتمية أسهل كثيرا، و هي تعطي:

$$\frac{\partial(\log_e L)}{\partial\beta_0} = \frac{1}{\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial(\log_e L)}{\partial\beta_1} = \frac{1}{\sigma^2} \sum X_i (Y_i - \beta_0 - \beta_1 X_i)$$

$$\frac{\partial(log_e L)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_i - \beta_0 - \beta_1 X_i)^2$$

نجعل الآن هذه المشتقات الجزئية مساوية للصفر، ونضع b_0 ، b_1 و $\hat{\sigma}^2$ بدلا من β_0 ، β_1 و σ^2، ونحصل بعد بعض التبسيطات على:

$$\sum (Y_i - b_0 - b_1 X_i) = 0 \qquad (2.30a)$$

$$\sum X_i (Y_i - b_0 - b_1 X_i) = 0 \qquad (2.30b)$$

$$\frac{\sum (Y_i - b_0 - b_1 X_i)^2}{n} = \hat{\sigma}^2 \qquad (2.30c)$$

تتطابق العلاقتان (2.30a) و (2.30b) مع معادلتي المربعات الدنيا الناظميتين المعطاتين سابقا في (2.9)، أما (2.30c) فهو مقدر σ^2 المنحاز المعطى سابقا في(2.27).

مسائل

(٢-١) بالعودة إلى مثال حجم المبيعات في صفحة (٢٧). إفترض أن عدد الوحدات المباعة مقاسة بدقة و لكن كثيرا ما تقع أخطاء كتابية في تحديد المبيعات بالدولار. هل تبقى العلاقة بين عدد الوحدات المباعة و المبيع بالدولار علاقة دالية؟ ناقش ذلك.

(٢-٢) يدفع أعضاء منتجع صحي رسوم عضوية سنوية قدرها 300$ بالإضافة إلى 2$ لكل زيارة إلى المنتجع. لنرمز بـ Y لإجمالي التكلفة السنوية للعضو بالدولارات و بـ X لعدد زيارات العضو خلال السنة. عبّر عن العلاقة بين X و Y رياضيا. هل هي دالية أم علاقة إحصائية؟

(٢-٣) توضح التجربة على نوع معين من البلاستيك أن هناك علاقة بين صلابة المواد المشكلة من البلاستيك (Y) (مقاسة بوحدات برنيل)، والوقت المنصرم منذ انتهاء عملية التشكيل (X). اقتُرح دراسة هذه العلاقة باستخدام تحليل الانحدار. واعترض أحد المشاركين في المناقشة، مشيرا إلى أن تصلب البلاستيك "نتاج عملية كيميائية طبيعية لا تترك أي شيء للمصادفة، و لذلك يجب أن تكون العلاقة رياضية وتحليل الإنحدار غير مناسب". ما هو تقويمك لهذا الإعتراض؟

(٢-٤) في الجدول (٢-١)، كان لدورتي الإنتاج 1 و 8 نفس حجم الدفعة X ولكنهما تختلفان في ساعات العمل Y. ما هي السّمة التي يوضحها ذلك لنموذج الانحدار في (2.1)؟

(٢-٥) كتب أحد الطلاب، عندما طُلب منه أن يعرض نموذج الانحدار الخطي البسيط، ما يلي $E\{Y_i\} = \beta_0 + \beta_1 X_i + \varepsilon_i$. هل توافق على ذلك؟

(٢-٦) لنعتبر نموذج الانحدار بخطأ طبيعي (2.25). ولنفترض أن قيم المعالم هي $\beta_0 = 200$ ، $\beta_1 = 5.0$ و $\sigma = 4$.

أ – ارسم نموذج الانحدار هذا بالطريقة المبينة في الشكل (٢-٧). ووضح توزيعات Y الموافقة للقيم $X = 10$، 20 و 40.

ب– فسر معاني المعالم β_0 و β_1، مفترضا أن مجال النموذج يغطي $X = 0$.

(٢-٧) في تمرين محاكاة، طبق نموذج انحدار (2.1) بـ $\beta_0 = 100$ ، $\beta_1 = 20$ و $\sigma^2 = 25$. ونريد الحصول على مشاهدة لـ Y من أجل $X = 5$.

أ – هل تستطيع أن تعطي بالضبط احتمال أن تقع Y بين 195 و 205؟ وضّح.

ب– إذا كان نموذج الإنحدار بخطأ طبيعي قابلا للتطبيق. هل يمكنك الآن إعطاء الاحتمال المضبوط لوقوع Y بين 195 و 205؟ و إذا كان الأمر كذلك فما هو هذا الاحتمال؟

(٢-٨) في الشكل (٢-٧)، لنفترض أننا حصلنا على مشاهدة أخرى لـ Y عند $X = 45$ هل سيبقى $E\{Y\}$ من أجل هذه المشاهدة الجديدة 104؟ و هل ستبقى قيمة Y لهذه المشاهدة الجديدة 108؟

(٢-٩) أعلن طالب محاسبة بحماس "إن الانحدار أداة قوية جدا. إذ يمكننا عزل التكاليف الثابتة والمتغيرة بتوفيق نموذج انحدار خطي، حتى عندما لا نمتلك بيانات عن الدفعات الصغيرة". ناقش.

(٢-١٠) درست محللة في تعاونية كبيرة العلاقة بين الرواتب السنوية الحالية (Y) والعمر (X) لـ 46 مبرمج حاسب آلي مستخدَمين حاليا في الشركة. واستنتجت أن العلاقة منحنية، تصل الذروة عند 47 سنة. هل يعني هذا أن رواتب المبرمجين تزداد حتى العمر 47 سنة و من ثم تنقص؟ وضح.

(٢-١١) دالة الانحدار التي تربط إنتاجية المستخدم بعد أخذ برنامج تدريبي (Y) وإنتاجيته قبل البرنامج التدريبي (X) هي $E\{Y\} = 20 + 0.95X$ حيث تمتد X من 40 إلى 100. استنتج مراقب أن برنامج التدريب لا يرفع الإنتاجية في المتوسط، لأن β_1 ليس أكثر من 1.0. علّق.

(٢-١٢) إشارة إلى المسألة (٢-٣)، يراد دراسة صلابة البلاستيك من أجل أربع فترات مختلفة انقضت منذ انتهاء عملية التشكيل (المعالجات). و تتوافر للدراسة ست عشرة عجنة (وحدات تجريبية). ويراد تخصيص كل معالجة لأربع وحدات تجريبية بصورة عشوائية. استخدم جدول الأرقام العشوائية، أو مولّد الأعداد العشوائية. للقيام بتخصيص عشوائي مناسب.

(٢-١٣) يُراد دراسة تأثير خمسة مستويات للجرعة في تصميم تام العشوائية، وتتوافر 20 وحدة تجريبية. ويُخصّص كل مستوى جرعة لأربع وحدات تجريبية نختارها عشوائيا. استخدم جدول الأرقام العشوائية، أو مولد الأعداد العشوائية، للقيام بتخصيص عشوائي مناسب.

(٢-١٤) قوّم العبارة التالية: "من أجل المشروعية التامة لتطبيق طريقة المربعات الدنيا يجب أن يكون توزيع Y طبيعيا".

(٢-١٥) يصرّح شخص أنه يمكن تقدير b_0 و b_1 في دالة الانحدار التوفيقية (2.12) بطريقة المربعات الدنيا. علّق.

(٢-١٦) وفقا للمعادلة (2.17) فإن $\Sigma e_i = 0$ عند توفيق نموذج الانحدار (2.1) لمجموعة من n مشاهدة مستخدمين طريقة المربعات الدنيا. هل يكون $\Sigma \varepsilon_i = 0$ أيضا؟ علّق.

(٢-١٧) **المعدل التراكمي.** طبقّ مدير القبول في كلية صغيرة اختبار دخول مصمم حديثا على عشرين طالبا اختيروا بصورة عشوائية من الطلاب الجدد في السنة الأولى وذلك في دراسة لتحديد ما إذا كان يمكن التنبؤ بمعدل الطالب التراكمي (GPR) في نهاية السنة الأولى (Y) بدءا من درجة اختيار الدخول (X). وفيما يلي نتائج الدراسة. افترض أن نموذج الانحدار من المرتبة الأولى (2.1) نموذج مناسب.

10	9	8	7	6	5	4	3	2	1	i
4.3	4.7	5.2	6.0	6.2	4.5	3.9	4.7	4.8	5.5	X_i
1.6	2.8	2.6	3.4	3.7	2.5	1.9	3.0	2.3	3.1	Y_i

20	19	18	17	16	15	14	13	12	11	i
4.7	4.1	5.9	5.0	4.3	4.6	6.3	5.0	5.4	4.9	X_i
1.5	2.2	3.8	2.0	1.4	1.8	3.2	2.3	2.9	2.0	Y_i

و فيما يلي ملخص لنتائج حسابية:

$$\Sigma Y_i^2 = 134.84, \Sigma X_i^2 = 205.12, \Sigma Y_i = 50.0, \Sigma X_i = 100.0, \Sigma X_i Y_i = 257.66$$

أ – أوجد مقدرات المربعات الدنيا لـ β_0 و β_1 واعرض دالة الانحدار المقدّرة.

ب– ارسم دالة الانحدار المقدرة و البيانات. هل تبدو دالة الانحدار المقدرة ملائمة للبيانات؟

جـ– أوجد تقديرا نقطيا للمعدل التراكمي للسنة الأولى لطلاب كانت درجتهم في اختبار الدخول $X = 5.0$.

د – ما هو التقدير النقطي لمتغير متوسط الاستجابة عندما تزداد درجة اختبار القبول نقطة واحدة؟

(٢-١٨) **صيانة الحاسبات.** تبيع تعاونية ترايسيتي للأجهزة المكتبية حاسبة يدوية مستوردة بموجب امتياز و تقوم بصيانة وقائية وخدمة إصلاح لهذه الحاسبة.

وجمعت البيانات أدناه من 18 طلبا حديثا مـن مستخدمي الحاسبات للقيام بخدمة صيانة وقائية روتينية ؛ و لكل طلب يمثل X عدد الآلات المخدومة وY عدد الدقائق الإجمالي التي استغرقها فني الصيانة. افترض أن نموذج الانحدار من المرتبة الأولى (2.1) مناسب.

i	1	2	3	4	5	6	7	8	9
X_i	7	6	5	1	5	4	7	3	4
Y_i	97	86	78	10	75	62	101	39	53

i	10	11	12	13	14	15	16	17	18
X_i	2	8	5	2	5	7	1	4	5
Y_i	33	118	65	25	71	105	17	49	68

و فيما يلي ملخص لنتائج حسابية:

$$\Sigma(Y_i - \overline{Y})^2 = 16.504 \ , \ \Sigma X_i = 81 \ , \ \Sigma Y_i = 1.152$$

$$\Sigma(X_i - \overline{X})^2 = 74.5, \ \Sigma(X_i - \overline{X})(Y_i - \overline{Y}) = 1.098$$

أ – أوجد دالة الانحدار المقدّرة.

ب– ارسم دالة الانحدار المقدرة والبيانات. ما مدى جودة توفيق دالة الانحدار المقدرة للبيانات؟

جـ – فسر b_0 في دالة الانحدار التي قدرتها. هل تعطي b_0 هنا أية معلومات مُفيدة؟ أوضح.

د – أوجد تقديرا نقطيا لمتوسط زمن الخدمة عند خدمة $X = 5$ آلات.

(٢–١٩) **الانكسار في الشحنات الجوية.** تُشحن مادة مما يُستخدم في الأبحاث الاحيائية و الطبية إلى المستخدمين جوا، و ذلك في صناديق تحتوي كل منها 1000 أنبولة. جُمعت البيانات أدناه، والتي تناولت عشر شحنات، عن عدد المرات (X) التي يحوّل فيها الصندوق من طائرة إلى أخرى خلال خط سير الشحنة، و(Y) عدد الأنبولات التي وجدت عند وصولها مكسورة. افترض أن نموذج الانحدار من الرتبة الأولى (2.1) مناسب.

i	1	2	3	4	5	6	7	8	9	10
X_i	1	0	2	0	3	1	0	1	2	0
Y_i	16	9	17	12	22	13	8	15	19	11

أ - أوجد دالة الانحدار المقدرة. وارسم دالة الانحدار المقدرة والبيانات. هل يبدو هنا أن دالة الانحدار الخطية تعطي توفيقا جيدا.

ب - أوجد تقديرا نقطيا للقيمة المتوقعة لعدد الأنبولات المكسورة عندما يكون عدد التحويلات $X=1$.

جـ - قدر الزيادة في القيمة المتوقعة لعدد الأنبولات المكسورة عندما يوجد تحويلان و ذلك بالمقارنة مع تحويل واحد.

د - تحقّقْ من أن خط الانحدار التوفيقي يمر من النقطة $(\bar{X},\bar{Y})$.

(٢-٢٠) **صلابة البلاستيك.** إشارة إلى المسألتين (٢-٣) و(٢-١٢)، صُنعت ست عشرة عجنة من البلاستيك، و شُكّلت وحدة اختبار واحدة من كل عجنة. وخُصصَت كل وحدة اختبار إلى أحد أربعة مستويات زمنية محددة سلفا، وقيست الصلابة بعد انقضاء الوقت المخصص. و النتائج مبينة أدناه: X الوقت المنصرم بالساعات و Y الصلابة مقاسة بوحدات برينل. افترض أن نموذج الانحدار من المرتبة الأولى (2.1) مناسب.

i	1	2	3	4	5	6	7	8
X_i	16	16	16	16	24	24	24	24
Y_i	199	205	196	200	218	220	215	223

i	9	10	11	12	13	14	15	16
X_i	32	32	32	32	40	40	40	40
Y_i	237	234	235	230	250	248	253	246

أ - أوجد دالة الانحدار المقدرة وارسم دالة الانحدار المقدرة و البيانات. هل يبدو هنا أن دالة انحدار خطية تعطي توفيقا جيدا؟

ب - أوجد تقديرا نقطيا لمتوسط الصلابة عندما $X = 40$ ساعة.

جـ - أوجد تقديرا نقطيا للتغير في متوسط الصلابة عندما تزداد X بساعة واحدة.

(٢-٢١) بالعودة إلى **مسألة المعدل التراكمي** (٢-١٧).

أ- أوجد الرواسب e_i. هل مجموعها صفر بما يتفق مع (2.17)؟

ب- قدر σ^2 و σ. بأي وحدات يُعبر عن σ ؟

(٢-٢٢) عد إلى المسألة (٢-١٨) الخاصة **بصيانة الحاسبات**.

أ - أوجد الرواسب e_i ومجموع مربعات الرواسب $\sum e_i^2$. ما هي العلاقة هنا بين مجموع مربعات الرواسب و الكمية Q في (2.8) ؟

ب- احصل على تقديرات نقطية لـ σ^2 وσ. بأي وحدات يُعبّر عن σ؟

(٢-٢٣) عد إلى المسألة (٢-١٩) الخاصة بالانكسار في شحنات جوية.

أ - أوجد الراسب للمشاهدة الأولى. ما هي علاقته بـ ε_1 ؟

ب- أحسب $\sum e_i^2$ وMSE . ما هي المعلمة التي يقدِّرها MSE ؟

(٢-٢٤) عد إلى مسألة (٢-٢٠) الخاصة **بصلابة البلاستيك**.

أ - أوجد الرواسب e_i. هل تَجمع إلى الصفر بما يتفق مع (2.17) ؟

ب- قدر σ^2 و σ. بأي وحدات يُعبر عن σ ؟

(٢-٢٥) **كتلة العضلة**. يُتوقع أن تقل كتلة عضلات الشخص مع العمر. ولتقصّي هذه العلاقة عند النساء، اختار باحث تغذية أربعة نساء عشوائيا من كل شريحة عمرية من 10 سنوات تبدأ بالعمر 40 وتنتهي بالعمر 79. وفيما يلي النتيجة؛ X العمر وY قياس كتلة العضلة. افترض أن نموذج الانحدار من المرتبة الأولى (2.1) مناسب.

8	7	6	5	4	3	2	1	i
56	68	73	56	67	43	64	71	X_i
80	78	73	87	68	100	91	82	Y_i

16	15	14	13	12	11	10	9	i
78	49	53	45	58	45	65	76	X_i
77	105	100	97	76	116	84	65	Y_i

أ - أوجد دالة الانحدار المقدّرة و ارسم الدالة المقدَّرة و البيانات. هل تبدو دالة الانحدار الخطية توفيقا جيدا؟ هل يدعم رسمك الاعتقاد بأن كتلة العضلة تتناقص مع العمر.

ب- أوجد ما يلي: (١) تقديرا نقطيا للاختلاف في متوسط كتلة العضلة لامرأتين يختلف عمراهما بسنة واحدة (٢) تقديرا نقطيا لمتوسط كتلة العضلة لامرأة عمرها X=60 سنة. (٣) قيمة الراسب للمشاهدة الثامنة،(٤) تقديرا نقطيا لِـ σ^2

(٢-٢٦) **معدل السرقات.** جمع مختص في علم الجريمة خلال دراسته للعلاقة بين الكثافة السكانية و معدل الجريمة في مدن أمريكية متوسطة الحجم، البيانات التالية من عينة عشوائية من 16 مدينة؛ X يمثل الكثافة السكانية في المدينة (عدد الأشخاص لكل وحدة مساحة) و Y معدل الجريمة خلال السنة الماضية (عدد السرقات لكل 100,000 شخص). افترض أن نموذج الانحدار من المرتبة الأولى (2.1) مناسب.

8	7	6	5	4	3	2	1	i
82	60	56	78	54	75	49	59	X_i
189	208	197	215	192	195	180	209	Y_i

16	15	14	13	12	11	10	9	i
70	89	65	47	94	88	83	69	X_i
204	200	186	205	212	214	201	213	Y_i

أ ـ أوجد دالة الانحدار المقدرة. ارسم دالة الانحدار المقدرة والبيانات. هل يبدو هنا أن دالة الانحدار الخطي تعطي توفيقا جيدا؟ ناقش.

ب ـ أوجد تقديرات نقطية لما يلي: (١) الاختلاف في متوسط معدل السرقة في مدن تختلف بمقدار الواحد في الكثافة السكانية. (٢) متوسط معدل السرقات خلال السنة الماضية في مدن كثافة سكانها $X = 60$. (٣) ε_{10}؛ (٤) σ^2.

تمارين

(٢-٢٧) بالعودة إلى نماذج الانحدار (2.1). لنفرض أن $X = 0$ ضمن مجال النموذج. إذا كانت $\beta_0 = 0$ بحيث يكون النموذج $Y_i = \beta_1 X_i + \varepsilon_i$ فماذا يترتب على ذلك بالنسبة لنموذج الانحدار؟ كيف ترسم دالة الانحدار بيانيا؟

(٢-٢٨) بالعودة إلى نموذج الانحدار (2.1). ماذا يعني في نموذج الانحدار كون $\beta_1 = 0$ وبالتالي كون النموذج $Y_i = \beta_0 + \varepsilon_i$؟ كيف يبدو الرسم البياني لدالة الانحدار.

(٢-٢٩) بالعودة إلى المسألة (٢-٢٠) **صلابة البلاستيك**. افرض أن وحدة اختبارية قد شُكلت من عجنة واحدة من البلاستيك وقيست صلابة هذه الوحدة عند 16 نقطة مختلفة من الزمن. هل لا يزال حد الخطأ في نموذج الانحدار هذا يعكس الآثار نفسها كما في التجربة الموصوفة سابقا؟ هل تتوقع أن تكون حدود الأخطاء للنقاط المختلفة في الزمن غير مرتبطة؟ ناقش.

(٢-٣٠) استنبط تعبيرا لـ b_1 في (2.10a) من المعادلات الناظمية في (2.9).

(٢-٣١) (يحتاج إلى التفاضل والتكامل) بالإشارة إلى نموذج الانحدار $Y_i = \beta_0 + \varepsilon_i$ في تمرين (٢-٢٨). استنبط مقدار المربعات الدنيا لـ β_0 في هذا النموذج.

(٢-٣٢) أثبت أن مقدر المربعات الدنيا لِـ β_0 المحسوب في تمرين (٢-٣١) غير منحاز.

(٢-٣٣) أثبت النتيجة في (2.18) ـ أن مجموع المشاهدات Y هو نفسه مجموع القيم التوفيقية.

(٢-٣٤) أثبت النتيجة في (2.20) ـ أن مجموع الرواسب الموزونة بالقيم التوفيقية يساوي صفرا.

(٢-٣٥) أثبت النتيجة الخاصة بـ SSE في (2.24a).

(٢-٣٦) بالعودة إلى الجدول (2.1) لمثال شركة وستوود. عندما سئل شخص لتقديم تقدير نقطي لمتوسط ساعات العمل لدورات من 60 قطعة، أعطى الشخص 131.7 كتقدير، لأن هذا هو متوسط ساعات العمل في الدورات الثلاث من الحجم 60 في الدراسة. وصرح ناقد أن طريقة هذا الشخص تُهمل معظم البيانات في الدراسة لأنها تجاهلت المشاهدات التي يختلف حجم الدفعة فيها عن 60. علّق.

(٢-٣٧) في مسألة **تكسر الشحنات** (٢-٩) كانت تقديرات المربعات الدنيا هي $b_0 = 10.20$، $b_1 = 400$ و $\sum e_i^2 = 17.60$. احسب معيار المربعات الدنيا في (2.8) المقابل للتقديرات (١) $b_0 = 9$ و$b_1 = 3$ ؛ (٢) $b_0 = 11$ و $b_1 = 5$.
هل المعيار Q أكبر لهذه التقديرات منه في تقديرات المربعات الدنيا؟

(٢-٣٨) حصلنا على مشاهدتين لـ Y عند كل من ثلاث مستويات لـ X وبالتحديد $X = 5$ ، $X = 10$ و $X = 15$.

أ - أثبت أن خط انحدار المربعات الدنيا التوفيقي للنقاط الثلاث $(5, \bar{Y}_1)$ ، $(10, \bar{Y}_2)$ و $(15, \bar{Y}_3)$ حيث $\bar{Y}_1$ ، $\bar{Y}_2$ و $\bar{Y}_3$ ترمز لمتوسطات المشاهدات Y عند مستويات X الثلاثة، متطابق مع خط انحدار المربعات الدنيات التوفيقي للحالات الست الأصلية.

ب - في هذه الدراسة، هل يمكن تقدير تباين حد الخطأ بدون توفيق خط انحدار؟ وضح.

(٢-٣٩) في مثال **شركة وستوود**، تقع مشاهدات Y عند، $X = 20$ و $X = 85$ على خط الانحدار التوفيقي مباشرة (جدول (٢-٣)وشكل (٢- ٢) ب) . إذا حُذفت هاتان المشاهدتان، هل يتغير خط انحدار المربعات الدنيا التوفيقي للحالات الثماني المتبقية؟ توضيح: ما هي مساهمة المشاهدتين في معيار المربعات الدنيا Q في (2.8)؟

(٢-٤٠) (يحتاج إلى التفاضل والتكامل). عُد إلى نموذج الانحدار في التمرين (٢-٢٧)، ثم:

أ - أوجد مقدر المربعات الدنيا لـ β_1.

ب - افرض أن حدود الأخطاء ε_i مستقلة، $N(0, \sigma^2)$ و σ^2 غير معروف. اكتب دالة الإمكانية العظمى لمشاهدات العينة Y وعددها n، وأوجد مقدر الإمكانية العظمى لِـ β_1. هل هو مطابق لمقدر المربعات الدنيا؟

جـ - أثبت أن مقدر الإمكانية العظمى لـ β_1 غير منحاز.

(٢-٤١) تجد أدناه عدد ألواح طباعة لمخطوطة (X) والتكلفة بالدولارات لتصحيح الأخطاء المطبعية (Y) في عينة عشوائية من طلبات استلمتها حديثا شركة متخصصة في المخطوطات التقنية. افرض أن نموذج الانحدار $Y_i = \beta_1 X_i + \varepsilon_i$ ، بحدود أخطاء طبيعية ومستقلة وتباينها $\sigma^2 = 16$، هو النموذج المناسب.

6	5	4	3	2	1	i
30	25	14	4	12	7	X_i
450	446	250	75	213	128	Y_i

أ ـ اكتب دالة الإمكانية لمشاهدات Y الست حيث $\sigma^2 = 16$.

ب ـ احسب دالة الإمكانية من أجل $\beta_1 = 17$، $\beta_1 = 18$ و$\beta_1 = 19$. لأي قيمة من قيم β_1 هذه تكون الإمكانية أكبر؟

جـ ـ إن مقدر الإمكانية العظمى هو $b_1 = \sum X_i Y_i / \sum X_i^2$ أوجد تقدير الإمكانية العظمى. هل تتسق نتائجك في (ب) مع هذا المقدر؟

مشاريع

(٢-٤٢) بالعودة إلى بيانات **SMSA** في الملحق (ب -٢). يتوقع وجود صلة بين عدد الأطباء العاملين في $(Y)SMSA$ وبين عدد السكان الكلي ومساحة المنطقة والدخل الإجمالي للفرد. افترض أن نموذج الانحدار من المرتبة الأولى (2.1) مناسب لكل من المتغيرات المستقلة الثلاثة.

أ ـ احدر عدد الأطباء العاملين على كل متغير بدوره من المتغيرات المستقلة الثلاثة واكتب دوال الانحدار المقدرة.

ب ـ ارسم دوال الانحدار المقدرة الثلاث والبيانات برسوم منفصلة. هل يبدو أن علاقة الانحدار الخطية تعطي توفيقا جيدا لكل من المتغيرات المستقلة الثلاثة؟

جـ ـ احسب MSE لكل من المتغيرات الثلاثة. أي متغير مستقل يقود إلى أقل تغير حول خط الانحدار التوفيقي؟

(٢-٤٣) بالعودة إلى مجموعة بيانات **SMSA** في الملحق (ب -٢).

أ ـ لكل منطقة جغرافية، احدر (Y) عدد الجرائم الخطرة في $SMSA$ على عدد السكان الكلي (X) افترض أن نموذج الانحدار من المرتبة الأولى مناسب لكل منطقة. اكتب دوال الانحدار المقدرة.

ب ـ هل دوال الانحدار المقدرة متشابهة للمناطق الأربع؟ ناقش.

جـ ـ احسب MSE لكل منطقة. هل تجد أن المتغير حول خط الانحدار التوفيقي يبقى نفسه تقريبا في المناطق الأربع؟ ناقش.

(٢-٤٤) بالعودة إلى مجموع بيانات **SENIC** في الملحق (ب - ١).

يتوقع أن يرتبط متوسط طول الإقامة في المستشفى مع خطورة الإصابة، والتسهيلات والخدمات المتوفرة، ونسبة الأشعة السينية الروتينية للصدر. افترض أن نموذج الانحدار من المرتبة الأولى مناسب لكل من المتغيرات المستقلة الثلاثة.

أ ـ احدر متوسط طول الإقامة على كل من المتغيرات المستقلة الثلاثة. اكتب دوال الانحدار المقدرة.

ب ـ ارسم دوال، الانحدار المقدرة الثلاث والبيانات برسوم منفصلة. هل يبدو أن العلاقة الخطية تمثل توفيقا جيدا لكل من المتغيرات المستقلة الثلاثة؟

جـ ـ احسب *MSE* لكل من المتغيرات المستقلة الثلاثة. أي متغير مستقل يقود إلى أقل تغير حول خط الانحدار التوفيقي؟

(٢ـ٤٥) بالعودة إلى مجموعة بيانات ***SENIC*** في الملحق (ب ـ ١).

أ ـ لكل منطقة جغرافية، احدر طول الإقامة في المستشفى (Y) على خطورة الإصابة (X). افرض أن نموذج الانحدار من المرتبة الأولى (2.1) مناسب لكل منطقة. اكتب دوال الانحدار المقدرة.

ب ـ هل تتشابه دوال الانحدار المقدرة للمناطق الأربع؟ ناقش.

جـ ـ احسب *MSE* لكل منطقة. هل يبقى التغير حول خط الانحدارالتوفيقي نفسه تقريبا للمناطق الأربع؟ ناقش.

استقراءات في تحليل الانحدار

في هذا الفصل، نتابع أولا استقراءات حول معلمتي الانحدار β_0 و β_1 معتبرين كلا من التقدير بفترة لهاتين المعلمتين والاختبارات حولهما. ومن ثم نناقش تقدير فترة للمتوسط $E\{Y\}$ للتوزيع الاحتمالي لـ Y من أجل X معطاة، وفترات تنبؤ لمشاهدة Y جديدة، من أجل X معطاة. وأخيرا نتابع تحليل الانحدار بأسلوب تحليل التباين، وطريقة الاختبار الخطي العام، ومقاييس وصفية للارتباط.

خلال هذا الفصل وما تبقى من الجزء I. وما لم نذكر خلاف ذلك، نفترض انطباق نموذج انحدار طبيعي الأخطاء (2.25). وهذا النموذج هو:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (3.1)$$

حيث:

β_0 و β_1 معلمتان

X_i ثوابت معروفة

ε_i مستقلة و $N(0, \sigma^2)$

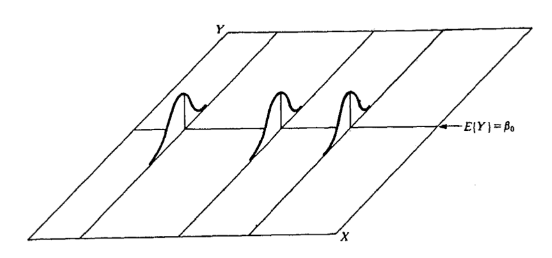
(٣-١) استقراءات حول β_1

كثيرا ما نهتم بالقيام باستقراءات حول β_1، ميل خط الانحدار في النموذج (3.1). فمثلا، قد يرغب محلل أبحاث تسويقية أن يدرس العلاقة بين المبيعات (Y) ونفقات الإعلان (X)، الحصول على تقدير فترة لـ β_1 لأنه يوفر معلومات عن متوسط كمية المبيعات الإضافية بالدولار، الناتجة عن إضافة دولار واحد في نفقات الإعلان.

كما يهتم، أحيانا، باختبارات حول β_1، وخصوصا تلك الاختبارات من الشكل:

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

شكل (٣-١) نموذج انحدار (3.1) عندما يكون $\beta_1 = 0$



وسبب الاهتمام في اختبار ما إذا كانت $\beta_1 = 0$ أم لا هو أن $\beta_1=0$ توضح عدم وجود علاقة خطية بين Y و X. ويوضح الشكل (٣-١) الحالة التي تكون فيها $\beta_1=0$ لنموذج الانحدار طبيعي الخطأ (3.1). ونلاحظ أن خط الانحدار أفقي وبالتالي فإن متوسطات التوزيعات الاحتمالية لـ Y متساوية، وبالتحديد:

$$E\{Y\} = \beta_0 + (0)X = \beta_0$$

وحيث إن نماذج الانحدار (3.1) تفترض توزيعات احتمالية طبيعية لـ Y بتباين ثابت وأن المتوسطات متساوية عندما $\beta_1=0$، فنجد أن التوزيعات الاحتمالية لـY متطابقة عندما $\beta_1=0$. ويوضح الشكل (٣-١) ذلك. وهكذا، لايؤدي $\beta_1 = 0$ في نموذج الانحدار (3.1) إلى عدم وجود علاقة خطية بين Y و X فقط، وإنما يؤدي كذلك إلى عدم وجود أي علاقة من أي نوع بين Y و X لأن التوزيعات الاحتمالية لـ Y متطابقة عند كل مستويات X.

ونحتاج، قبل مزيد من مناقشة الاستقراءات حول β_1، إلى دراسة توزيع المعاينة لـb_1، المقدّر النقطي لـ β_1.

توزيع معاينة b_1

أُعطي المقدِّر النقطي b_1 في العلاقة ($2.10\,a$) كمايلي:

$$b_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} \tag{3.2}$$

ويشير توزيع المعاينة لـ b_1 إلى القيم المختلفة لـ b_1 التي يمكن الحصول عليها عند تكرار المعاينة مع بقاء مستويات المتغير المستقل X ثابتة من عينة إلى عينة.

ومن أجل نموذج الانحدار (3.1)، يكون توزيع المعاينة لـ b_1 طبيعيا بمتوسط وتباين:

$$E\{b_1\} = \beta_0 \tag{3.3a}$$

$$\sigma^2\,\{b_1\} = \frac{\sigma^2}{\sum(X_i - \overline{X})^2} \tag{3.3b}$$

ولتبيان ذلك، نحتاج إلى معرفة أن b_1 تركيب خطي في المشاهدات Y_I.

b_1 كتركيب خطي في الـ Y_I. يمكن تبيان أنه يمكن التعبير عن b_1 كما هي معرفة في (3.2) كما يلي:

$$b_1 = \sum k_i Y \tag{3.4}$$

حيث:

$$k_i = \frac{X_i - \overline{X}}{\sum(X_i - \overline{X})^2} \tag{3.4a}$$

لاحظ أن k_i كميات ثابتة لأن X_i مثبتة. وهكذا فإن b_1 تركيب خطي في Y_i حيث المعاملات ليست إلا دالة في المقادير المثبتة X_I.

وللمعاملات k_i عدد من الخصائص المهمة التي ستستخدم فيما بعد:

$$\sum k_i = 0 \tag{3.5}$$

$$\sum k_i X_i = 1 \tag{3.6}$$

$$\sum k_i^2 = \frac{1}{\sum(X_i - \overline{X})^2} \tag{3.7}$$

تعليقات

١ ـ لتبيان أن b_1 تركيب خطي في الـ K_I، نثبت أولا:

$$\sum(X_i - \overline{X})(Y_i - \overline{Y}) = \sum(X_i - \overline{X})Y_i \tag{3.8}$$

وينتج هذا لأن:

$$\sum(X_i-\overline{X})(Y_i-\overline{Y})=\sum(X_i-\overline{X})Y_i-\sum(X_i-\overline{X})\overline{Y}$$

ولكن $\sum(X_i-\overline{X})\overline{Y}=\overline{Y}\sum(X_i-\overline{X})=0$ لأن $\sum(X_i-\overline{X})=0$. وهكذا تتحقق (3.8).

ونعبر الآن عن b_1 مستخدمين (3.8) و (3.4 a) كما يلي:

$$b_1=\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{\sum(X_i-\overline{X})^2}=\frac{\sum(X_i-\overline{X})Y_i}{\sum(X_i-\overline{X})^2}=\sum k_iY_i$$

٢- إثبات خصائص k_i مباشر. فعلى سبيل المثال، تتبع الخاصية (3.5) لأن:

$$\sum k_i=\sum\left[\frac{X_i-\overline{X}}{\sum(X_i-\overline{X})^2}\right]=\frac{\sum(X_i-\overline{X})}{\sum(X_i-\overline{X})^2}=\frac{0}{\sum(X_i-\overline{X})^2}=0$$

وبالمثل تتبع الخاصية (3.7) لأن:

$$\sum k_i^2=\sum\left[\frac{(X_i-\overline{X})}{\sum(X_i-\overline{X})^2}\right]^2=\frac{1}{\left[\sum(X_i-\overline{X})^2\right]^2}\sum(X_i-\overline{X})^2$$

$$=\frac{1}{\sum(X_i-\overline{X})^2}$$

الطبيعية. نعود الآن إلى توزيع المعاينة لـ b_1 في حالة النموذج طبيعي الخطأ (3.1). وتنتج طبيعية توزيع المعاينة لـ b_1 مباشرة من حقيقة أن b_1 تركيب خطي في الـ Y_i. والـ Y_i تتوزع طبيعيا ومستقلة طبقا للنموذج (3.1). وتعرض النظرية (1.37) أن التركيب الخطي في متغيرات عشوائية طبيعية مستقلة يتوزع طبيعيا.

متوسط. ومن السهل تبيان عدم انحياز المقدر النقطي b_1 المذكور سابقا في نظرية جاوس - ماركوف (2.11):

$$E\{b_1\}=E\left\{\sum k_iY_i\right\}=\sum k_iE\{Y_i\}=\sum k_i(\beta_0+\beta_1X_i)$$

$$=\beta_0\sum k_i+\beta_1\sum k_iX_i$$

وبالاستناد إلى (3.5) و (3.6) نجد بالتالي أن:

$$E\{b_1\}=\beta_1$$

تباين. يمكن اشتقاق تباين b_1 بسهولة. ونحتاج فقط إلى تذكر أن الـ Y_i متغيرات

عشوائية مستقلة، تباين كل منها σ^2، وأن k_i ثوابت. من ثم نحصل من (1.28) على:

$$\sigma^2\{b_1\} = \sigma^2\{\sum k_i Y_i\} = \sum k_i^2 \sigma^2\{Y_i\}$$
$$= \sum k_i^2 \sigma^2 = \sigma^2 \sum k_i^2$$
$$= \sigma^2 \frac{1}{\sum (X_i - \bar{X})^2}$$

والخطوة الأخيرة تتبع من (3.7).

تباين مقدّر. يمكننا تقدير تباين توزيع المعاينة لـ b_1:

$$\sigma^2\{b_1\} = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

وذلك بأن نستبدل بالمعلمة σ^2 مقدرا غير منحاز لها، ونقصد MSE:

$$s^2\{b_1\} = \frac{MSE}{\sum (X_i - \bar{X})^2} = \frac{MSE}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} \tag{3.9}$$

فالمقدر النقطي $s^2\{b_1\}$ مقدر غير منحاز لِ $\sigma^2\{b_1\}$. وبأخذ الجذر التربيعي نحصل على $s\{b_1\}$، المقدر النقطي لِ $\sigma\{b_1\}$.

ملاحظة

عرضنا في النظرية (2.11) أن لِ b_1 تباينا أصغريا بين كافة المقدرات الخطية غير المنحازة من الشكل:

$$\hat{\beta}_1 = \sum c_i Y_i$$

حيث الـ c_i ثوابت كيفية. وسوف نثبت هذا الآن. بما أن $\hat{\beta}_1$ يجب أن يكون غير منحاز فلابد أن يتحقق التالي:

$$E\{\hat{\beta}_1\} = E\{\sum c_i Y_i\} = \sum c_i E\{Y_i\} = \beta_1$$

والآن لدينا من (2.2) $E\{Y_i\} = \beta_0 + \beta_1 X_i$ ، وبذلك يصبح الشرط أعلاه:

$$E\{\hat{\beta}_1\} = \sum c_i(\beta_0 + \beta_1 X_i) = \beta_0 \sum c_i + \beta_1 \sum c_i X_i = \beta_1$$

ولتحقق شرط عدم الانحياز يجب أن تخضع الـ c_i للقيود:

$$\sum c_i = 0 \qquad \sum c_i X_i = 1$$

ومن (1.28) يكون تباين $\hat{\beta}_1$ الآن:

$$\sigma^2\{\hat{\beta}_1\} = \sum c_i^2 \sigma^2\{Y_i\} = \sigma^2 \sum c_i^2$$

ليكن $c_i = k_i + d_i$ تعريفا، حيث الـ K_i هي ثوابت المربعات الدنيا في (3.4a) والـ d_i ثوابت كيفية. وبالتالي يمكن كتابة:

$$\sigma^2\{\hat{\beta}_1\} = \sigma^2 \sum c_i^2 = \sigma^2 \sum (k_i + d_i)^2 = \sigma^2 \left(\sum k_i^2 + \sum d_i^2 + 2\sum k_i d_i \right)$$

نعلم من برهاننا أعلاه أن $\sigma^2 \sum K_i^2 = \sigma^2\{b_1\}$. إضافة إلى أن $\sum k_i d_i = 0$. بسبب القيود المفروضة أعلاه على الـ k_i والـ c_i :

$$\begin{aligned}
\sum k_i d_i &= \sum k_i (c_i - k_i) \\
&= \sum c_i k_i - \sum k_i^2 \\
&= \sum c_i \left[\frac{X_i - \overline{X}}{\sum (X_i - \overline{X})^2} \right] - \frac{1}{\sum (X_i - \overline{X})^2} \\
&= \frac{\sum c_i X_i - \overline{X} \sum c_i}{\sum (X_i - \overline{X})^2} - \frac{1}{\sum (X_i - \overline{X})^2} = 0
\end{aligned}$$

وبالتالي لدينا:

$$\sigma^2\{\hat{\beta}_1\} = \sigma^2\{b_1\} + \sigma^2 \sum d_i^2$$

لاحظ أن القيمة الأصغر لـ $\sum d_i^2$ هي الصفر. وبالتالي يكون تباين $\hat{\beta}_1$ أصغريا عندما يكون $\sum d_i^2 = 0$. ولكن لا يمكن حدوث هذا إلا عندما تكون جميع المقادير d_i مساوية للصفر، مما يتضمن كون $c_i \equiv k_i$.

وهكذا، يمتلك مقدِّر المربعات الدنيا، من بين كل المقدرات الخطية غير المنحازة، تباينا أصغريا.

توزيع المعاينة لـ $(b_1 - \beta_1)/s\{b_1\}$

بما أن b_1 يتوزع طبيعيا، فنعلم أن الإحصاءة المعيارية $(b_1 - \beta_1)/\sigma\{b_1\}$ متغير طبيعي معياري. وبالطبع نحتاج عادة إلى تقدير $\sigma\{b_1\}$ بـ $s\{b_1\}$ ، وبالتالي نهتم بتوزيع الإحصاءة المعيارية $(b_1 - \beta_1)/s\{b_1\}$. وتعرض نظرية مهمة في الإحصاء مايلي:

(3.10) $\dfrac{b_1 - \beta_1}{s\{b_1\}}$ يتوزع في نماذج الانحدار (3.1) وفق توزيع $t(n-2)$.

وبالبداهة، ينبغي أن لا تكون هذه النتيجة غير متوقعة. إذ نعلم أنه إذا جاءت المشاهدات Y_i من مجتمع طبيعي فإن $(\bar{Y}-\mu)/s\{\bar{Y}\}$ يتبع توزيع t بـ $n-1$ من درجات الحرية. والمقدر b_1، شأنه شأن $\bar{Y}$، تركيب خطي في المشاهدات Y_i. ويعود سبب الاختلاف في درجات الحرية إلى الحاجة إلى تقدير المعلمتين (β_1 و β_0) في نموذج الانحدار. وبالتالي نفقد هنا درجتي حرية.

ملاحظة

يمكننا إثبات أن $(b_1-\beta_1)/s\{b_1\}$ يتوزع وفق t بـ $n-2$ من درجات الحرية استنادا إلى النظرية التالية:

(3.11) من أجل نماذج الانحدار (3.1)، يتوزع SSE/σ^2 وفق χ^2 بعدد $n-2$ درجة حرية وبصورة مستقلة عن b_1 و b_0.

أولا، دعنا نعيد كتابة $(b_1-\beta_1)/s\{b_1\}$ كما يلي:

$$\frac{b_1-\beta_1}{\sigma\{b_1\}} \div \frac{s\{b_1\}}{\sigma\{b_1\}}$$

فيكون البسط متغيرا طبيعيا معياريا z. ويمكن رؤية طبيعية المقام بأن نعتبر أولا:

$$\frac{s^2\{b_1\}}{\sigma^2\{b_1\}} = \frac{\dfrac{MSE}{\sum(X_i-\bar{X})^2}}{\dfrac{\sigma^2}{\sum(X_i-\bar{X})^2}} = \frac{MSE}{\sigma^2} = \frac{\dfrac{SSE}{n-2}}{\sigma^2}$$

$$= \frac{SSE}{\sigma^2(n-2)} \sim \frac{\chi^2(n-2)}{n-2}$$

حيث يعني الرمز ~ "موزع وفق". وتتبع الخطوة الأخيرة من (3.11). وبالتالي لدينا:

$$\frac{b_1-\beta_1}{s\{b_1\}} \sim \frac{z}{\sqrt{\dfrac{\chi^2(n-2)}{n-2}}}$$

ولكن z و χ^2 مستقلان استنادا إلى النظرية (3.11)، ذلك لأن z دالة في b_1 و b_1 مستقل عن $SSE/\sigma^2 \sim \chi^2$. وهكذا نستنتج من تعريف (1.41)، أن:

$$\frac{b_1-\beta_1}{s\{b_1\}} \sim t(n-2)$$

وتضعنا هذه النتيجة في موقع نستطيع معه القيام باستقراءات سهلة حول β_1.

فترة الثقة لـ β_1

بما أن $(b_1 - \beta_1)/s\{b_1\}$ يتبع التوزيع t، فيمكننا صياغة العبارة الاحتمالية التالية:

$$P\{t(\alpha/2;\ n\text{-}2) \le (b_1-\beta_1)/\ s\{b_1\} \le t(1-\alpha/2;\ n-2)\} = 1 - \alpha \qquad (3.12)$$

ويرمز $t(\alpha/2; n-2)$ هنا إلى المئين $100(\alpha/2)$ للتوزيع t بـ $n-2$ درجة حرية. ويتبع من تناظر التوزيع t أن:

$$t(\alpha/2;\ n-2) = -t(1-\alpha/2;\ n-2) \qquad (3.13)$$

وبإعادة ترتيب المتباينات في (3.12)، واستخدام (3.13)، نحصل على:

$$P\{b_1 - t(1-\alpha/2;\ n-2)\ s\{b_1\} \le \beta_1 \le b_1 + t(1-\alpha/2;\ n-2)s\{b_1\}\} = 1-\alpha \qquad (3.14)$$

وبما أن (3.14) تتحقق من أجل جميع قيم β_1 الممكنة فإن α - 1 حدي ثقة لـ β_1 هما:

$$b_1 \pm t(1-\alpha/2;\ n-2)s\{b_1\} \qquad (3.15)$$

مثال. دعنا نعود إلى مثال حجم الدفعة لشركة وستوود في الفصل الثاني. ترغب الإدارة تقديرا لـ β_1 بـ 95 بالمائة معامل ثقة. ونلخّص في الجدول (٣-١) النتائج اللازمة والتي سبق الحصول عليها. إذ نحتاج أولا إلى الحصول على $s\{b_1\}$:

$$s^2\{b_1\} = \frac{MSE}{\sum(X_i-\bar{X})^2} = \frac{7.5}{3,400} = 0.002206$$

$$s\{b_1\} = 0.04697$$

ويتطلب 95 بالمائة معامل ثقة، المقدار $t(0.975;8)$. ومن جدول (أ ـــ ٢) في الملحق أ، نجد $t(0.975;8) = 2.306$. ومن (3.15) تكون الـ 95 بالمائة فترة ثقة:

$$2.0-2.306(0.04697) \le \beta_1 \le 2.0 + 2.306(0.04697)$$
$$1.89 \le \beta_1 \le 2.11$$

وهكذا، وبمعامل ثقة 0.95، نقدّر أن متوسط عدد ساعات العمل تزداد بما يتراوح بين 1.89 و2.11 لكل زيادة في حجم الدفعة مقدارها جزء واحد.

الجدول (٣ ـ ١) نتائج لمثال شركة وستوود حصلنا عليها في الفصل الثاني.

$n = 10$	$\bar{X} = 50$
$b_0 = 10.0$	$b_1 = 2.0$
$\hat{Y} = 10.0 + 2.0X$	$SSE = 60$
$\sum X_i^2 = 28{,}400$	$MSE = 7.5$
$\sum X_i^2 - \frac{(\sum X_i)^2}{n} = \sum (X_i - \bar{X})^2 = 3{,}400$	
$\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = 6{,}800$	
$\sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = \sum (Y_i - \bar{Y})^2 = 13{,}660$	

ملاحظة

لاحظنا، في الفصل الثاني، أن مجال نموذج الانحدار مقيد، عادة، لفترة ما من قيم المتغير المستقل. ومن المهم تذكّر هذا، على وجه الخصوص، عند استخدام تقديرات الميل β_1. في مثالنا عن حجم الدفعة، يبدو أن نموذج الانحدار الخطي مناسب لحجوم دفعات بين 20 و80، وهو مدى المتغير المستقل في الماضي القريب. وقد لا يكون من المنطقي استخدام تقديرات الميل للاستقراءات عن تأثير حجم الدفعة على عدد ساعات العمل بعيدا خارج هذا المدى حيث إن علاقة الانحدار قد لا تكون هنا، خطية.

اختبارات حول β_1

بما أن $(b_1 - \beta_1)/s\{b_1\}$ يتوزع وفقا لـ t بـ $n - 2$ درجة حرية، يمكن القيام، باختبارات حول β_1 مستخدمين توزيع t بالطريقة المعتادة.

مثال ١ . اختبار ثنائي الجانب. يرغب محلل تكاليف في شركة وستوود في اختبار ما إذا كانت توجد صلة خطية بين ساعات العمل وحجم الدفعة أم لا، وذلك باستخدام نموذج الانحدار (3.1). والبديلان عندئذٍ هما :

$$H_0: \beta_1 = 0 \qquad H_a: \beta_1 \neq 0 \tag{3.16}$$

وإذا رغب المحلل بضبط مخاطرة الخطأ من النوع الأول عند 0.05، فباستطاعته، في

الحقيقة، استنتاج H_a حال العودة إلى الـ 95 بالمائة فترة ثقة لـ β_1 الـتي أقمناهـا سـابقا، لأن الفترة لا تتضمن الصفر.

ويستند اختبار واضح للبديلين (3.16) على إحصاءة الاختبار:

$$t^* = \frac{b_1}{s\{b_1\}} \tag{3.17}$$

وتكون قاعدة القرار مع إحصاءة الاختبار هذه، عند ضبط مستوى المعنوية عند αهي:

$$\begin{aligned} &\text{استنتج } H_0 \text{ إذا كان } |t^*| \leq t(1 - \alpha/2; n-2) \\ &\text{استنتج } H_a \text{ إذا كان } |t^*| > t(1 - \alpha/2; n-2) \end{aligned} \tag{3.17a}$$

وفي مثال شركة وستوود، حيث $\alpha = 0.05$، $b_1 = 2.0$ و $s\{b_1\}$=0.04697 نجـد $t(0.975;8) = 2.306$. وهكذا تكون قاعدة القرار لاختبار البديلين (3.16)

استنتج H_0 إذا كان $|t^*| \leq 2.306$

استنتج H_a إذا كان $|t^*| > 2.306$

وبما أن $|t^*| = |2.0/0.04697| = 42.58 > 2.306$ فنسـتنتج H_a. أي $\beta_1 \neq 0$ أو هنـاك صلة خطية بين ساعات العمل وحجم الدفعة.

نحصل على القيمة - P لنتيجة العينة بإيجاد الاحتمـال $P\{t(8) > t^* = 42.58\}$. ومن الجدول (أـ٢) نرى أن الاحتمال أقل من 0.0005 وفي الحقيقة، يمكن إثبـات أنهـا تقريبا صفر وسيُرمز لها بـ 0^+. وهكذا، فإن القيمـة -P ثنائيـة الجـانب هـي $2(0^+)=0^+$ وحيث إن القيمة -P ثنائية الجانب أقل من مستوى المعنوية المحـدد $\alpha = 0.05$ فيمكننـا مباشرة استنتاج H_a.

مثال٢. اختبار وحيد الجانب. لو أن المحلل رغب في اختبار ما إذا كان β_1 موجبـا أم لا، مع ضبط مستوى المعنوية عند $\alpha = 0.05$ لكان البديلان:

$$H_0: \beta_1 \leq 0$$
$$H_a: \beta_1 > 0$$

ولكانت قاعدة القرار المبنية على إحصاءة الاختبار (3.17):

استنتج H_0 إذا كان $t^* \leq t(1 - \alpha; n-2)$

استنتج H_a إذا كان $t^* > t(1 - \alpha; n-2)$

من أجل $\alpha = 0.05$، نحتاج إلى $t(0.95;8) = 1.860$. وبما أن $t^* = 42.58 > 1.860$ فنستنتج H_a، أي أن β_1 موجب.

وكان بالإمكان الوصول مباشرة إلى هذا القرار نفسه من القيمة - P وحيدة الجانب والتي لوحظ أنها 0^+ في المثال١. وبما أن القيمة -P هذه أقل من0.5، فنستنتج H_a.

تعليقات

١ - تنشر كثير من حزم الحاسب والنشرات العلمية، عادة، القيمة - P مع قيمة إحصاءة الاختبار. وبهذه الطريقة يمكن إجراء اختبار عند أي مستوى معنوية مرغوب، وذلك بمقارنة القيمة -P مع المستوى المحدد α. وينبغي على مستخدمي حزم الحاسب أخذ الحذر والتحقق مما إذا كانت القيمة -P المعطاة وحيدة الجانب أم ثنائية الجانب.

٢ - نرغب أحيانا في اختبار ما إذا كان β_1 يساوي قيمة محددة β_{10} غير الصفر أم لا، والتي ربما تكون تاريخيا قيمة قياسية، أو القيمة لتجربة مشابهة، أو مواصفة هندسية. وفي اختبار كهذا، تكون الإحصاءة المناسبة:

$$t^* = \frac{b_1 - \beta_{10}}{s\{b_1\}} \tag{3.18}$$

وقاعدة القرار المستخدمة للبديلين:

$$H_0: \beta_1 = \beta_{10}$$
$$H_a: \beta_1 \neq \beta_{10}$$

هي القاعدة (3.17a) نفسها ولكنها، الآن مبنية على t^* المعرف في (3.18).

لاحظ أن إحصاءة الاختبار (3.18) تُختزل إلى إحصاءة الاختبار (3.17) عندما ينطوي الاختبار على $H_0: \beta_1 = \beta_{10} = 0$.

(٣-٢) استقراءات حول β_0

كما لوحظ في الفصل الثاني، فليس هناك إلا مناسبات غير متواترة نرغب فيها باستقراءات حول β_0، مقطوع خط الانحدار. ويحدث هذا عندما يتضمن مجال النموذج القيمة $X = 0$.

توزيع المعاينة لـ b_0

أُعطي المقدِّر النقطي b_0 في (2.10b) كما يلي:

$$b_0 = \overline{Y} - b_1 \overline{X} \tag{3.19}$$

ويعود توزيع المعاينة لـ b_0 إلى الاختلاف في قيم b_0 التي نحصل عليها عند تكرار المعاينة مع بقاء مستويات المتغير المستقل X ثابتة من عينة إلى أخرى.

(3.20) في نموذج الانحدار (3.1)، يكون توزيع المعاينة لـ b_0 طبيعيا بمتوسط وتباين.

$$E\{b_0\} = \beta_0 \tag{3.20a}$$

$$\sigma^2\{b_0\} = \sigma^2 \frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2} = \sigma^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum(X_i - \overline{X})^2}\right] \tag{3.20b}$$

وتنتج طبيعية توزيع المعيانة لـ b_0 لأن b_0، مثله مثل b_1، تركيب خطي في المشاهدات Y_i. ويمكن الحصول على نتائج المتوسط والتباين لتوزيع المعاينة لـ b_0 بطريقة مشابهة لتلك المستخدمة من أجل b_1.

ونحصل على مقدّر لـ $\sigma^2\{b_0\}$ بأن نستبدل بـ σ^2 تقديرها النقطي MSE

$$s^2\{b_0\} = MSE \frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2} = MSE\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum(X_i - \overline{X})^2}\right] \tag{3.21}$$

ويكون الجذر التربيعي $s\{b_0\}$ مقدرا لـ $\sigma\{b_0\}$.

توزيع المعاينة لـ $(b_0 - \beta_0)/s\{b_0\}$.

توجد نظرية حول b_0 مشابهة للنظرية (3.10) حول b_1 وتعرض:

(3.22) في النموذج (3.1) يتوزع $\dfrac{b_0 - \beta_0}{s\{b_0\}}$ وفق $t(n-2)$.

وبالتالي يمكن صياغة فترات ثقة لـ β_0 واختبارات حول β_0، بالطريقة المعتادة، مستخدمين التوزيع t.

فترة ثقة لـ β_0

نحصل على $1-\alpha$ حدي ثقة لـ β_0 بالطريقة نفسها التي اشتقت فيها سابقا من أجل β_1:

$$b_0 \pm t(1 - \alpha/2; n-2)s\{b_0\} \tag{3.23}$$

مثال. كما لوحظ سابقا لا يمكن مد مجال نموذج مثال شركة وستوود إلى حجوم دفعات $X = 0$. وبالتالي، قد لا تملك معلمة الانحدار β_0 معنى ذا مغزى هنا. ومع ذلك،

إذا رغبنا في 90 بالمائة فترة ثقة لـ β_0 فإننا نمضي بإيجاد $t(0.95;8)$ و $s\{b_0\}$ ومن الجدول (أ - ٢)، نجد $t(0.95;8) = 1.86$. وباستخدام النتائج السابقة الملخصة في الجدول (٣-١) نحصل من (3.21) على:

$$s^2\{b_0\} = MSE\frac{\sum X_i^2}{n\sum(X_i - \bar{X})^2} = 7.5\left[\frac{28,400}{10(3,400)}\right] = 6.26471$$

أو

$$s\{b_0\} = 2.50294$$

وبالتالي تكون الـ 95 بالمائة فترة ثقة لـ β_0 هي:

$$10.0 - 1.860\,(2.50294) \le \beta_0 \le 10.0 + 1.860(2.50294)$$
$$5.34 \le \beta_0 \le 14.66$$

ونحذّر مرة ثانية أنه ليس من الضروري أن تزودنا فترة الثقة هذه بأية معلومات ذات مغزى. وعلى سبيل المثال، ليس ضروريا أن تزودنا بمعلومات عن التكاليف التأسيسية لإنتاج دفعة من الأجزاء (التكاليف التي يتطلبها بناء عملية الإنتاج بغض النظر عن حجم الدفعة) لأننا لسنا متأكدين مما إذا كان نموذج الانحدار الخطي مناسبا عند توسيع مجال النموذج بحيث يشمل $X = 0$.

(٣ - ٣) بعض الاعتبارات عند القيام باستقراءات حول β_0 و β_1

تأثير الابتعاد عن الطبيعية

إذا لم تكن التوزيعات الاحتمالية لـ Y طبيعية ولكنها لا تبتعد ابتعادا لا يمكن إغفاله عن التوزيع الطبيعي، فإن توزيعي المعاينة لـ b_0 و b_1 يكونان طبيعيين تقريبا، وسيزودنا استخدام التوزيع t بمعاملات الثقة المحددة أو بمستوى المعنوية بصورة تقريبية. أما إذا ابتعدت توزيعات Y عن الطبيعي، فللمقدّرين b_0 و b_1 عموما خاصية التقارب إلى الطبيعي - يقترب توزيعهما من الطبيعي تحت شروط عامة جدا مع زيادة حجم العينة. وهكذا من أجل عينات كبيرة بما فيه الكفاية تبقى فترات الثقة وقواعد القرار المعطاة سابقا قابلة للتطبيق حتى لو ابتعدت التوزيعات الاحتمالية لـ Y عن الطبيعي. ومن أجل عينات كبيرة نستبدل، قيمة t بقيمة z للتوزيع الطبيعي المعياري.

تفسير معاملات الثقة ومخاطر الخطأ

بما أن نموذج الانحدار (3.1) يفترض أن الـ X_i ثوابت معروفة، فإن معاملات الثقة

ومخاطر الأخطاء تتخذ تفسيرا لها من خلال أخذ عينات متكررة تحافظ فيها المشاهدات على المستويات نفسها كما في العينة الملحوظة. وعلى سبيل المثال، أقمنا فترة ثقة لـ β_1 بمعامل ثقة 0.95 في مثال شركة وستوود. وتفسير هذا المعامل هو أنه إذا أخذنا العديد من العينات المستقلة مع بقاء مستويات X (أحجام الدفعات) في العينات المتكررة كما كانت في العينة الأولى، وأقمنا 95 بالمائة فترة ثقة لكل عينة، فإن 95 بالمائة من هذه الفترات سوف تحتوي القيمة الحقيقية لـ β_1.

المسافات بين مستويات X المتتالية

يوضح تأمل العلاقة بين (3.3b) و (3.20b) لتبايني b_0 و b_1 ، على الترتيب أنه من أجل n و σ^2 معروفين يتأثر هذان التباينان بالمسافات بين المستويات المتتالية لـ X في البيانات الملحوظة. فمثلا، كلما ازداد الانتشار في مستويات X، كبرت الكمية $\sum(X_i - \bar{X})^2$ وصغر تباين b_1 . وسوف نناقش في الفقرة (٥ـ٧) كيف تُحدَّد مواقع المشاهدات X في التجارب التي يمكن التحكم فيها بمسافات هذه المشاهدات بعضها عن بعض.

قوة الاختبار

يمكن الحصول على قوة اختباري β_0 و β_1 من الجدول (أ ــ ٥) في الملحق أ، والذي يحوي جداول بيانية لدالة قوة الاختبار t. اعتبر، مثلا مسألة القرار العامة:

$$H_0: \beta_1 = \beta_{10} \qquad (3.24)$$
$$H_a: \beta_1 \neq \beta_{10}$$

والتي استخدمت فيها إحصاءة الاختبار العامة (3.18):

$$t^* = \frac{b_1 - \beta_{10}}{s\{b_1\}} \qquad (3.24a)$$

وكانت قاعدة القرار لمستوى معنوية α هي:

استنتج H_0 إذا كان $|t^*| \leq t(1 - \alpha/2; n-2)$ (3.24b)

استنتج H_a إذا كان $|t^*| > t(1 - \alpha/2; n-2)$

فتكون قوة هذا الاختبار هي احتمال أن تقود قاعدة القرار إلى النتيجة H_a عندما تكون، في الواقع، صحيحة. وبالتحديد تُعطى القوة بـ:

$$\text{القوة} = P\{|t^*| > t(1 - \alpha/2; n-2) \mid \delta\} \qquad (3.25)$$

حيث δ مقياس اللامركزية ـ أي، كم تبعد قيمة β_0 الحقيقية عن β_{10}:

$$\delta = \frac{|\beta_1 - \beta_{10}|}{\sigma\{b_1\}} \tag{3.26}$$

يقدم الجدول (أ ـ ٥) قوى اختبار t ثنائي الجانب (كنسب مئوية) وذلك من أجل α=0.01 و 0.05= α لمختلف درجات الحرية df. ولتوضيح استخدام هذا الجدول، دعنا نرجع إلى مثال شركة وستوود حيث اختبرنا:

$$H_0: \beta_1 = \beta_{10} = 0$$
$$H_a: \beta_1 \neq \beta_{10} = 0$$

افترض أننا نرغب معرفة قوة الاختبار عندما تكون β_1= 0.25. فلتحقيق ذلك نحتاج إلى معرفة σ^2، تباين حدود الخطأ. افترض أن σ^2 = 10.0 وبذلك تكون $\sigma^2\{b_1\}$ في مثالنا:

$$\sigma^2\{b_1\} = \frac{\sigma^2}{\sum(X_i - \overline{X})^2} = \frac{10.0}{3,400} = 0.002941$$

أو $\sigma\{b_1\}$ =0.05423 وعندئذٍ $\delta = |0.25 - 0| \div 0.05423 = 4.6$. ندخل الرسم البياني وفق α = 0.05 (مستوى المعنوية المستخدم في الاختبار). ونرسم بصورة تقريبية وبالعين المجردة المنحنى الخاص بثمانية درجات حرية. وبقراءة للإحداثي الصادي الموافق لـ $\delta = 4.6$ نحصل تقريبا على 0.97 بالمائة. وهكذا إذا كانت β_1 = 0.25، فسيقودنا الاختبار إلى النتيجة Ha $(\beta_1 \neq 0)$ باحتمال 97. تقريبا. وبعبارة أخرى، إذا كانت β_1=.25 فسنكون شبه متأكدين من استنتاج أن هناك علاقة خطية بين ساعات العمل وحجم الدفعة.

ويمكن الحصول على قوة الاختبارات حول β_0 من الجدول (ا ـ ٥) بطريقة مشابهة تماما. ومن أجل اختبارات وحيدة الجانب ندخل الجدول (ا ـ ٥) على أساس أن نصف مستوى المعنوية المعروضة هناك هو مستوى المعنوية في الاختبار وحيد الجانب.

(٣ ـ ٤) تقدير الفترة لـ $E\{Y_h\}$

أحد الأهداف الرئيسية في تحليل الانحدار هو، عادة، تقدير متوسط توزيع واحد أو أكثر من توزيعات Y الاحتمالية. اعتبر، مثلا، دراسة العلاقة بين مستوى الأجر على أساس القطعة X وإنتاجية العامل Y. فقد يكون لمتوسط الإنتاجية عند مستويات عالية

ومتوسطة للأجر على أساس القطعة أهمية خاصة لأغراض تحليل الأرباح الناتجة عن زيادة في الأجر. وكمثال آخر، قد تهتم شركة وستوود بمتوسط الاستجابة (متوسط عدد ساعات العمل) لدفعات حجمها $X = 40$ قطعة، $X = 55$ قطعة و$X = 70$ قطعة، وذلك لأغراض اختيار أحجام الدفعات المناسبة للإنتاج.

لنرمز بـ X_h لمستوى X الذي نرغب بتقدير متوسط الاستجابة من أجله. وقد تكون X_h قيمة من قيم العينة : أو قد تكون قيمة أخرى للمتغير المستقل ضمن مجال النموذج. نرمز لمتوسط الاستجابة عند $X = X_h$ بـ $E\{Y_h\}$. وتعطينا العلاقة (2.12) المقدّر النقطي $\hat{Y}_n$ لـ $E\{Y_h\}$:

$$\hat{Y}_h = b_0 + b_1 X_h \qquad (3.27)$$

نستعرض الآن توزيع المعاينة لـ $\hat{Y}_h$.

توزيع المعاينة لـ $\hat{Y}_h$

يشير توزيع المعاينة لـ $\hat{Y}_h$، كما هو الحال في توزيعات المعاينة التي ناقشناها سابقا، إلى القيم المختلفة لـ $\hat{Y}_h$ التي تنتج عند تكرار إختيار عينات كل منها يحفظ مستويات المتغير المستقل X ثابتة، وحساب $\hat{Y}_h$ لكل عينة.

(3.28) ولنموذج الانحدار (3.1)، يكون توزيع المعاينة لـ $\hat{Y}_h$ طبيعيا بمتوسط وتباين:

$$E\{\hat{Y}_h\} = E\{Y_h\} \qquad (3.28a)$$

$$\sigma^2\{\hat{Y}_h\} = \sigma^2 \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] \qquad (3.28b)$$

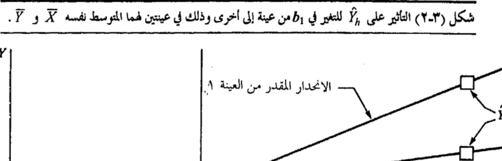
الطبيعية. تنتج طبيعية توزيع المعاينة لـ $\hat{Y}_n$ مباشرة من حقيقة أن $\hat{Y}_n$ تركيب خطي في المشاهدات Y_i.

المتوسط. لإثبات أن $\hat{Y}_n$ مقدّر غير منحاز لـ $E\{Y_h\}$، نمضي كالتالي:

$$\begin{aligned} E\{\hat{Y}_h\} &= E\{b_0 + b_1 X_h\} = E\{b_0\} + X_h E\{b_1\} \\ &= \beta_0 + \beta_1 X_h \end{aligned}$$

وذلك بالاستناد إلى (3.3a) و(2.20a).

شكل (٣-٢) التأثير على $\hat{Y}_h$ للتغير في b_1 من عينة إلى أخرى وذلك في عينتين لهما المتوسط نفسه $\bar{X}$ و $\bar{Y}$.



التباين. لاحظنا أن متغيرية توزيع المعاينة لـ $\hat{Y}_h$ تتأثر بمقدار بُعد X_h عن $\bar{X}$ من خلال الحد $(X_h - \bar{X})^2$. فكلما ابتعد X_h عن $\bar{X}$ كبرت الكمية $(X_h - \bar{X})^2$ مما يزيد في تباين $\hat{Y}_h$. ويعطي الشكل (٣-٢) توضيحا بديهيا لهذا التأثير، فهناك نعرض توفيقين لخطي انحدار على أساس عينتين تتفقان في مجموعة قيم X. ولنفترض أن خطي الانحدار يمران من النقطة $(\bar{X}, \bar{Y})$ نفسها، وذلك لعزل التأثير موضع الاهتمام، ونعني تأثير تغير الميل المقدَّر b_1 من عينة إلى عينة. لاحظ أنه عندما تكون X_1 قريبة من $\bar{X}$ تكون القيمتان التوفيقيتان $\hat{Y}_1$ لخطي انحدار العينتين قريبتين من بعضهما. ويكون الوضع مختلفا عند X_2 البعيدة عن $\bar{X}$. فهنا تختلف القيمتان التوفيقيتان $\hat{Y}_2$ كثيرا. وهكذا فإن تأثير تغير الميل b_1 من عينة إلى أخرى على $\hat{Y}_h$ هو تأثير أوضح بكثير عند مستويات X البعيدة عن $\bar{X}$ منه عند مستويات X القريبة من $\bar{X}$. وبالتالي يكون التغير في قيم $\hat{Y}_h$ من عينة إلى أخرى أكبر، في حالة X_h بعيدة عن المتوسط، منه في حالة X_h قريبة من المتوسط.

وعند التعويض بـ MSE عن σ^2 في (3.28b)، نحصل على $s^2\{\hat{Y}_h\}$، التباين المقدَّر لـ $\hat{Y}_h$:

$$s^2\{\hat{Y}_h\} = MSE\left[\frac{1}{n}+\frac{(X_h-\overline{X})^2}{\sum(X_i-\overline{X})^2}\right] \tag{3.29}$$

ويكون الانحراف المعياري المقدَّر لـ $\hat{Y}_h$، عندئذٍ، الجذر التربيعي لـ $s^2\{\hat{Y}_h\}$.

ملاحظة

لاستنباط $\sigma^2\{\hat{Y}_h\}$ نبين أولا أن b_1 و $\overline{Y}$ غير مرتبطين وبالتالي فهمـا في نمـوذج الانحـدار (3.1) مستقلان:

$$\sigma\{\overline{Y},b_1\}=0 \tag{3.30}$$

حيث يرمز $\sigma\{\overline{Y},b_1\}$ للتغاير بين $\overline{Y}$ و b_1. ونبدأ بالتعاريف:

$$\overline{Y}=\sum\left(\frac{1}{n}\right)Y_i$$

$$b_1=\sum k_iY_i$$

حيث k_i هي كما عرفناها في (3.4a). والآن نستخدم النظرية (1.29) مـع $a_i=\frac{1}{n}$ و $c_i=k_i$ متذكرين أن الـ Y_i متغيرات عشوائية مستقلة:

$$\sigma\{\overline{Y},b_1\}=\sum\left(\frac{1}{n}\right)k_i\sigma^2\{Y_i\}=\frac{\sigma^2}{n}\sum k_i$$

ولكن نعلم من (3.5) أن $\sum K_i=0$ وبالتالي فإن التغاير 0.

والآن نحن جاهزون لإيجاد تباين $\hat{Y}_h$. وسوف نستخدم المقـدِّر في الصيغـة البديلـة (2.15):

$$\sigma^2\{\hat{Y}_h\}=\sigma^2\left\{\overline{Y}+b_1(X_h-\overline{X})\right\}$$

وبما أن $\overline{Y}$، b_1 مستقلان و X_h ، $\overline{X}$ ثابتان، فنجد:

$$\sigma^2\{\hat{Y}_h\}=\sigma^2\{\overline{Y}\}+(X_h-\overline{X})^2\sigma^2\{b_1\}$$

والآن $\sigma^2\{b_1\}$ يُعطى في (3.3b) و:

$$\sigma^2\{\overline{Y}\}=\frac{\sigma^2\{Y_i\}}{n}=\frac{\sigma^2}{n}$$

وبالتالي:

$$\sigma^2\{\hat{Y}_h\} = \frac{\sigma^2}{n} + (X_h - \bar{X})^2 \frac{\sigma^2}{\sum(X_i - \bar{X})^2}$$

والتي، تعطي، بعد شيء من ترتيب الحدود العلاقة (3.28b).

توزيع المعاينة لـ $\left(\hat{Y}_h - E\{Y_h\}\right)/s\{\hat{Y}_h\}$

حيث إننا واجهنا توزيع t في كل أنواع الاستقراءات حتى الآن في نموذج الانحدار (3.1) فينبغي أن لا يكون مفاجئا أن:

$$\frac{\hat{Y}_h - E(Y_h)}{s\{\hat{Y}_h\}} \quad (3.31)$$

يتوزع وفقا لـ $t(n-2)$ في نموذج الانحدار (3.1).

وهكذا تجري جميع الاستقراءات حول $E\{Y_h\}$ بالطريقة المعتادة مع توزيع t. وسنوضح بناء فترات الثقة لأن استخدامها في التطبيقات يتكرر أكثر من الاختبارات.

فترة الثقة لـ $E\{Y_h\}$

تبنى فترة الثقة $E\{Y_h\}$ بالطريقة المعتادة باستخدام توزيع t كما هو موضح في النظرية (3.31). و $(1-\alpha)$ حدي ثقة هما:

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{\hat{Y}_h\} \quad (3.32)$$

مثال ١. بالعودة إلى مثال حجم الدفعة لشركة وستوود، دعنا نجده 90% فترة ثقة لـ $E\{Y_h\}$ عندما يكون حجم الدفعة $X_h = 55$ قطعة. وباستخدام النتائج السابقة في الجدول (٣-١) نجد التقدير النقطي $\hat{Y}_h$:

$$\hat{Y}_h = 10.0 + 2.0(55) = 120$$

والآن نحتاج إلى حساب الانحراف المعياري المقدّر $s\{\hat{Y}_h\}$. وباستخدام (3.29) نحصل على:

$$s^2\{\hat{Y}_h\} = 7.5\left[\frac{1}{10} + \frac{(55-50)^2}{3{,}400}\right] = 0.80515$$

وهكذا يكون:

$$s\{\hat{Y}_h\} = 0.89730$$

ومن أجل 90% فترة ثقة، نحتاج إلى $t(0.95;8) = 1.860$. وبالتالي تكون فترة ثقتنا

بمعامل ثقة 0.90 هي باستخدام (3.32):

$$120 - 1.860(0.89730) \le E\{Y_h\} \le 120 + 1.860(0.89730)$$
$$118.3 \le E\{Y_h\} \le 121.7$$

ونستنتج بمعامل ثقة 0.90 أن متوسط عدد ساعات العمل المطلوبة عند إنتاج دفعات من 55 قطعة يقع بين 118.3 و 121.7.

مثال ٢. افترض أن شركة وستوود ترغب في تقدير $E\{Y_h\}$ عندما $X_h = 80$ قطعة وذلك بـ 90% فترة ثقة. نحتاج هنا إلى :

$$\hat{Y}_h = 10.0 + 2.0(80) = 170$$
$$s^2\{\hat{Y}_h\} = 7.5\left[\frac{1}{10} + \frac{(80-50)^2}{3,400}\right] = 2.73529$$

$$s\{\hat{Y}_h\} = 1.65387$$
$$t(0.95;8) = 1.860$$

وهكذا تكون الـ 90% فترة ثقة:

$$170\text{-}1.860(1.65387) \le E\{Y_h\} \le 120 + 1.860(1.65387)$$
$$166.9 \le E\{Y_h\} \le 173.1$$

لاحظ أن فترة الثقة هنا أوسع بعض الشيء من تلك في المثال ١، لأن مستوى الـ X_h هنا ($X_h = 80$) أكثر بعدا عن المتوسط $\overline{X} = 50$ من مستوى الـ X_h في المثال ١، ($X_h = 55$).

تعليقات

١ - بما أن الـ X_i في نموذج الانحدار (3.1) ثوابت معلومة، فإن فترات الثقة ومخاطر الأخطاء في الاستقراءات حول متوسط الاستجابة تُفسر من خلال تصور أخذ عينات متكررة تحافظ فيها المشاهدات X على المستويات نفسها كما في العينة الملحوظة فعلا. وقد نوّهنا سابقا بهذه النقطة، فيما يتعلق باستقراءات حول β_0 و β_1.

٢ - نرى من العلاقة ($3.28b$) أنه من أجل نتائج معطاة لعينة، يكون تباين $\hat{Y}_h$ أصغريا عندما يكون $X_h = \overline{X}$. ولذلك، ففي تجربة لتقدير متوسط الاستجابة عند مستوى معين X_h للمتغير المستقل، تكون دقة التقدير أعظمية إذا اتخذت (مع بقاء كل شيء آخر على حاله) المشاهدات في X مواقعها بحيث يكون $\overline{X} = X_h$.

٣ ـ عندما يكون حجم العينة كبيرا يمكن، في حدود الثقة (3.32) استبدال قيمة z الطبيعية المعيارية بالقيمة t، لأن توزيع t يتقارب إلى التوزيع الطبيعي المعياري مع ازدياد حجم العينة.

٤ ـ تنطبق العلاقات المعتادة بين فترات الثقة والاختبارات في استقراءات تتعلق بمتوسط الاستجابة. وهكذا يمكن الانتفاع بحدي الثقة ذات الجانبين (3.32) في الاختبارات ذات الجانبين المتعلقة بمتوسط الاستجابة عند X_h. ويمكن بصورة بديلة، وضع قاعدة قرار نظامية.

٥ ـ حدا الثقة (3.32) لمتوسط الاستجابة $E\{Y_h\}$ غير حسّاسين لابتعاد معتدل عن فرضية أن حدود الأخطاء تتوزع طبيعيا. في الواقع، لا تكون الحدود حسّاسة حتى لابتعاد شديد عن الطبيعية إذا كان حجم العينة كبيرا. وهذه الخاصية عند تقدير متوسط الاستجابة مرتبطة بخاصية عدم تأثر حدود الثقة لـ β_0 و β_1 بالابتعاد عن الطبيعية والتي لاحظناها سابقا.

٦ ـ ينطبق حدا الثقة (3.32) عندما نريد تقدير متوسط استجابة وحيد من العينة. ونناقش في الفصل الخامس كيفية العمل عندما يُراد تقدير عدّة متوسطات استجابة من العينة نفسها.

(٣ ـ ٥) التنبؤ بمشاهدة جديدة

نعتبر الآن التنبؤ بمشاهدة جديدة Y ، مقابلة لمستوى معين X للمتغير المستقل. وفي توضيح شركة وستوود، على سبيل المثال، فإن الدفعة القادمة المراد إنتاجها تتألف من 55 قطعة، وترغب الإدارة بالتنبؤ بعدد ساعات العمل اللازمة لهذه الدفعة بالذات. وكمثال آخر، قدَّر اقتصادي علاقة الانحدار بين مبيعات شركة وعدد الأشخاص الذين أعمارهم 16 سنة فما فوق، استنادا إلى بيانات من السنوات العشر الماضية. ومع توافر إسقاط سكاني موثوق لعدد الأشخاص الذين أعمارهم 16 سنة فأكثر في السنة القادمة، يرغب الاقتصادي التنبؤ بمبيعات الشركة للسنة القادمة.

ويُنظر للمشاهدة الجديدة Y على أنها محاولة جديدة مستقلة عن المشاهدات التي استند إليها تحليل الانحدار. وسنرمز لمستوى X في المحاولة الجديدة بـ X_h وللمشاهدة Y الجديدة بـ $Y_{h(new)}$. وبالطبع، نفترض أن نموذج الانحدار المعتمد، والقابل للتطبيق على بيانات العينة الأساسية يبقى مناسبا للمشاهدة الجديدة.

والتمييز أساسي بين تقدير متوسط الاستجابة $E\{Y_h\}$ الذي نوقش في الفقرة السابقة والتنبؤ باستجابة جديدة $Y_{h(new)}$ الذي نناقشه الآن. ففي الحالة السابقة نقدر متوسط توزيع Y. والآن، نتنبأ بنتيجة بمفردها مسحوبة من توزيع Y. وبالطبع فإن الغالبية العظمى من النتائج المفردة تنحرف عن متوسط الاستجابة. ويجب أن يكون هذا متاحا في عملية التنبؤ بـ $Y_{h(new)}$.

فترة التنبؤ عند معرفة المعالم

لتوضيح طبيعة فترة التنبؤ بمشاهدة جديدة $Y_{h(new)}$ بطريقة بسيطة قدر الإمكان، سنفترض أولا أن كل معالم الانحدار معروفة. وسنلغي لاحقا هذا الافتراض ونقوم بتعديلات مناسبة.

افترض أن شركة وستوود تخطط لإنتاج دفعة من $X_h = 40$ قطعة في أسابيع قليلة، ومعروف أن المعالم ذات العلاقة في نموذج الانحدار هي:

$$\beta_0 = 9.5 \qquad \beta_1 = 2.1$$
$$E\{Y\} = 9.5 + 2.1X$$
$$\sigma^2 = 10.0$$

وهكذا نجد من أجل $X_h = 40$ قطعة:

$$E\{Y_h\} = 9.5 + 2.1(40) = 93.5$$

ويوضح الشكل (٣-٣) التوزيع الاحتمالي لـ Y في حالة $X_h = 40$ قطعة. ومتوسطة $E\{Y_h\} = 93.5$، وانحرافه المعياري $\sigma = \sqrt{10.0} = 3.162$. وفضلا عن ذلك فالتوزيع طبيعي بما يتفق مع نموذج الانحدار (3.1).

افترض أننا في صدد التنبؤ بأن عدد ساعات العمل للدفعة $X_h = 40$ قطعة القادمة سيكون بين:

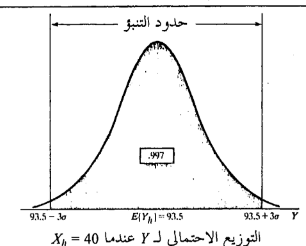
$$E\{Y_h\} \pm 3\sigma$$
$$93.5 \pm 3(3.162)$$

وبذلك تكون فترة التنبؤ:

$$84.0 < Y_{h(new)} < 103.0$$

وبما أن 99.7 بالمائة من مساحة التوزيع الاحتمالي تقع ضمن ثلاثة انحرافات معيارية عن المتوسط زيادة أو نقصا فإنه وباحتمال 0.997، ستعطي فترة التنبؤ هذه تنبؤا صحيحا لشوط الإنتاج القادم لـ 40 قطعة.

شكل (٣-٣) التنبؤ بـ $Y_{h(new)}$ عندما تكون المعالم معروفة.



التوزيع الاحتمالي لـ Y عندما $X_h = 40$

وعلى هذا تكون الفكرة الأساسية لفترة التنبؤ هي اختيار مدى في توزيع Y تقع فيه أغلب المشاهدات، والادّعاء بأن المشاهدة القادمة سوف تقع في هذا المدى. وتعتمد فائدة فترة التنبؤ، كما هو الحال دائما، على عرض الفترة وحاجة المستَخْدِم للدقة.

وعموما، عندما تكون معالم الانحدار معروفة فإن الـ $1 - \alpha$ حدي تنبؤ لـ $Y_{h(new)}$ تكون:

$$E\{Y_h\} \pm z(1 - \alpha/2)\sigma \quad (3.33)$$

وبتمركز الحدين حول $E\{Y_h\}$، نحصل على أضيق فترة متلائمة مع الاحتمال المحدد للتنبؤ الصحيح.

فترة الثقة لـ $Y_{h(new)}$ عندما تكون المعالم غير معروفة

يجب تقدير معالم الانحدار عندما تكون غير معروفة. ويُقدَّر متوسط توزيع Y كالمعتاد بـ $\hat{Y}_h$ ويُقدَّر تباين توزيع Y بـ MSE. ولا نستطيع، في جميع الأحوال، أن

نستخدم، ببساطة حدي التنبؤ (3.33) بعد تعويض المعالم بالمقدّرات النقطية المقابلة لها. وبالبداهة يوضح الشكل (٣-٤) سبب ذلك. إذ يُعرض هناك توزيعان احتماليان لـ Y، يقابلان الحدين الأعلى والأدنى لفترة الثقة الخاصة بـ $E\{Y_h\}$ وبعبارة أخرى يمكن أن يتموضع توزيع Y كأقصى توزيع مبين إلى اليسار، أو كالتوزيع الآخر المبين بعيدا إلى اليمين أو في أي مكان بينهما. وبما أننا لا نعلم المتوسط $E\{Y_h\}$ ولكننا فقط نقدره بفترة ثقة، فلا يمكننا التأكد من موقع توزيع Y.

ويوضّح الشكل (٣-٤) كذلك حدي التنبؤ لكل من توزيعي Y الاحتماليين المعروضين هناك. وبما أننا غير متأكدين من موقع توزيع Y، فمن الواضح أنه يجب أن تأخذ حدود تنبؤ $Y_{h(new)}$ في الحساب عنصرين، كما هو مبين في الشكل (٣-٤):

١- التغيرات في المواقع الممكنة لتوزيع Y.

٢- التغيرات ضمن توزيع Y الاحتمالي.

ونحصل على حدي تنبؤ لمشاهدة Y جديدة عند مستوى مُعطى X_h، باستخدام النظرية التالية:

(3.34) في نموذج الانحدار (3.1) يتوزع $\dfrac{\hat{Y}_h - Y}{s\left\{Y_{h(new)}\right\}}$ وفقا لـ$t(n-2)$.

لاحظ أن الإحصاءة المعيارية (3.34) تستخدم المقدّر النقطي $\hat{Y}_h$ في البسط بدلا من المتوسط الحقيقي $E\{Y_h\}$ لأن المتوسط الحقيقي غير معروف، ولا يمكن استخدامه للقيام بتنبؤ، وسوف نعرِّف بعد قليل الانحراف المعياري المقدّر $S\{Y_h(new)\}$ الموجود في مقام الإحصاءة المعيارية.

ومن النظرية (3.34) نجد بالطريقة المعتادة أن الـ $(1-\alpha)$ حدي تنبؤ لمشاهدة جديدة هما [قارن، مثلا، (3.34) مع (3.10) واربط $\hat{Y}_h$ مع b_1 و Y مع β_1]

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{Y_{h(new)}\} \tag{3.35}$$

ونحصل بسهولة على تباين البسط للإحصاءة المعيارية (3.34) وذلك بالاستفادة من استقلال المشاهدة الجديدة Y ومشاهدات العينة الأصلية التي بُنيت عليها $\hat{Y}_h$. وسنرمز إلى تباين البسط بـ $\sigma^2\{Y_{h(new)}\}$، ونحصل عليه باستخدام العلاقة (1.28b).

$$\sigma^2\{Y_{h(new)}\} = \sigma^2\{\hat{Y}_h - Y\} = \sigma^2\{\hat{Y}_h\} + \sigma^2\{Y\} = \sigma^2\{\hat{Y}_h\} + \sigma^2 \qquad (3.36)$$

لاحظ أن للتباين $\sigma^2\{Y_{h(new)}\}$ مركبتين:

١ - تباين توزيع المعاينة لـ $\hat{Y}_h$.

٢ - تباين توزيع Y عند $X = X_h$.

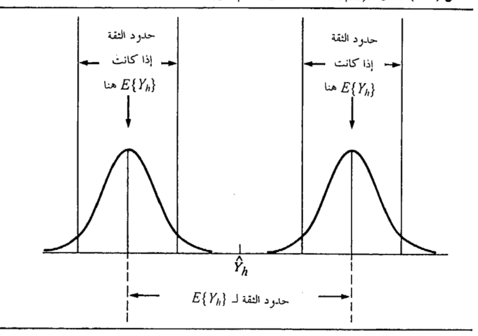
وكمقدّر غير منحاز لِـ $\sigma^2\{Y_{h(new)}\}$ نجد:

$$s^2\{Y_{h(new)}\} = s^2\{\hat{Y}_h\} + MSE \qquad (3.37)$$

ويمكن التعبير عنه مستخدمين (3.29) كالتالي:

$$s^2\{Y_{h(new)}\} = MSE\left[1 + \frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right] \qquad (3.37a)$$

شكل (٣-٤) التنبؤ بـ $Y_{h(new)}$ عندما لا تكون المعالم معروفة.



مثال. افترض أن شركة وستوود ترغب في تنبؤ عن عدد ساعات العمل التي يتطلبها شوط إنتاج قادم من الحجم 55 بـ 90 بالمائة فترة تنبؤ، وأن قيم المعالم غير معروفة. نحتاج إلى $t(0.95;8) = 1.860$. ولدينا من عمل سابق:

$$\hat{Y}_h = 120 \qquad s^2\{\hat{Y}_h\} = 0.80515$$

$$MSE = 7.5$$

وباستخدام (3.37) نحصل على:

$$s^2\{Y_{h(new)}\} = 0.80515 + 7.5 = 8.30515$$

وهكذا يكون:

$$s\{Y_{h(new)}\} = 2.88187$$

وبالتالي تكون 90 بالمائة فترة تنبؤ لـ $Y_{h(new)}$ من (3.35):

$$120 - 1.860(2.88187) \le Y_{h(new)} \le 120+1.860(2.88187)$$
$$114.6 \le Y_{h(new)} \le 125.4$$

وبمعامل ثقة 0.90 نتنبأ أن عدد ساعات العمل اللازمة لشوط الإنتاج القادم لـ 55 قطعة سوف تتراوح بين 114.6 و 125.4 ساعة.

تعليقات

١ - إن الـ 90 بالمائة فترة تنبؤ لـ $Y_{h(new)}$ التي حصلنا عليها آنفا أوسع من الـ 90 بالمائة فترة ثقة لـ $E\{Y_{55}\}$ التي حصلنا عليها في مثال ١ صفحة ٩٥ والسبب هو أننانواجه عند التنبؤ بمشاهدة جديدة كلا من التغيرات في $\hat{Y}_h$ من عينة إلى عينة بالإضافة إلى التغير ضمن التوزيع الاحتمالي لـY.

٢ - تشير العلاقة ($3.37a$) أن فترة التنبؤ تتسع مع ابتعاد X_hعن $\bar{X}$. والسبب في ذلك هو أن تقدير متوسط $\hat{Y}_h$ ، كما لوحظ سابقا، يكون أقل دقة كلما ابتعد مع موقع X_hعن $\bar{X}$.

٣ - خلافا لحدي الثقة (3.32) لمتوسط الاستجابة $E\{Y_h\}$ ، فإن حدي التنبؤ (3.35) حسّاسان للابتعاد عن طبيعية توزيع حدود الخطأ. ونناقش في الفصل الرابع طرقا تشخيصية لفحص طبيعية التوزيع الاحتمالي لحدود الخطأ، ونصف تدابير علاجية إذا كان الابتعاد عن الطبيعية كبيرا.

٤ - يشير معامل الثقة لحدي التنبؤ (3.35) إلى تكرار أخذ عينات ترتكز على المجموعة نفسها من قيم X، ثم حساب حدي التنبؤ لـ $Y_{h(new)}$ لكل عينة.

٥ - يعبر حد التنبؤ نفسيهما إلى استخدامات السيطرة الإحصائية ــ فلنفترض في مثال شركة وستوود أن شوط الإنتاج الجديد لـ 55 قطعة، والتي كان حدا التنبؤ

لحاجته من ساعات العمل هما 114.6 و 125.4 ساعة، قد تطلب في واقع الأمر 135 ساعة. فقد تتخذ الإدارة من هذا مؤشرا إلى إمكانية حدوث تغير في عملية الإنتاج، وربما ترغب الشروع في بحث عن السبب الذي يقف وراء ذلك.

٦ ـ عندما يكون حجم العينة كبيرا، يكون الحدان الأخيران داخل الأقواس في (3.37a) صغيرين بالمقارنة مع 1، الحد الأول بين القوسين. وبالطبع يكون التوزيع t، عندئذٍ، قريبا من التوزيع الطبيعي. وهكذا فإن الـ ($1 - \alpha$) حدي تنبؤ تقريبيين لـ $Y_{h(new)}$ عندما تكون n كبيرة هما:

$$\hat{Y}_h \pm z(1-\alpha/2)\sqrt{MSE} \tag{3.38}$$

٧ ـ يُطبق حدا التنبؤ (3.35) على تنبؤ وحيد يعتمد على بيانات العينة. ونناقش فيما يلي كيفية التنبؤ بمتوسط عدد من المشاهدات الجديدة عند مستوى معلوم X_h؛ وفي الفصل الخامس نتابع كيفية القيام بعدة تنبؤات عند مستويات X_h مختلفة.

٨ ـ تشبه فترات التنبؤ فترات الثقة. ومع ذلك فإنهما يختلفان من حيث المفهوم. فتمثل فترة الثقة استقراء حول معلمة، وهي فترة المقصود منها تغطية قيم المعلمة. ومن جهة أخرى فإن فترة التنبؤ هي عبارة حول القيمة التي سيأخذها متغير عشوائي.

التنبؤ بمتوسط m من المشاهدات الجديدة عن X_h معطاة

يرغب أحدنا أحيانا التنبؤ بمتوسط m من المشاهدات Y الجديدة عند مستوى معطى للمتغير المستقل. افترض أنه طُلب من شركة وستوود التقدم بعطاء في مناقصة تستدعي $m = 3$ من الأشواط الإنتاجية المستقلة لـ $X_h = 55$ قطعة وذلك خلال الأشهر القليلة القادمة. وترغب الإدارة التنبؤ بمتوسط ساعات العمل لكل شوط من الأشواط الثلاثة ومن ثَمَّ تحويل هذه إلى تنبؤ عن عدد ساعات العمل الكلية التي يتطلبها إتمام العقد.

وسنرمز بـ $\overline{Y}_{h(new)}$ لمتوسط قيمة Y التي سنتنبأ بها. ويمكن تبيان أن الـ ($1 - \alpha$) حدي تنبؤ المناسبين هما:

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{\overline{Y}_{h(new)}\} \tag{3.39}$$

حيث:

$$s^2\{\overline{Y}_{h(new)}\} = s^2\{\hat{Y}_h\} + \frac{MSE}{m} \qquad (3.39a)$$

أو بصورة مكافئة:

$$s^2\{\overline{Y}_{h(new)}\} = MSE\left[\frac{1}{m} + \frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right] \qquad (3.39b)$$

لاحظ من $(3.39a)$ أن للتباين $s^2\left\{\overline{Y}_{h(new)}\right\}$ مركبتين:

١- تباين توزيع معاينة $\hat{Y}_h$.

٢- تباين متوسط المشاهدات الـ m من توزيع Y الاحتمالي عند $X = X_h$.

مثال. في مثال شركة وستوود، دعنا نحسب الـ 90 بالمائة فترة تنبؤ لمتوسط عدد ساعات العمل $\overline{Y}_{h(new)}$ في ثلاثة أشواط إنتاجية جديدة، ولكل منها $X_h = 55$ قطعة. لدينا من عمل سابق:

$$\hat{Y}_h = 120 \qquad s^2\{\hat{Y}_h\} = .80515$$

$$MSE = 7.5 \qquad t(0.95;8) = 1.860$$

وبالتالي نجد:

$$s^2\{\overline{Y}_{h(new)}\} = 0.80515 + \frac{7.5}{3} = 3.30515$$

أو

$$s^2\{\overline{Y}_{h(new)}\} = 1.81801$$

وعندئذٍ تكون فترة التنبؤ لمتوسط عدد ساعات العمل للشوط الواحد:

$$120 - 1.860(1.81801) \le \overline{Y}_{h(new)} \le 120 + 1.860(1.81801)$$

$$116.6 \le \overline{Y}_{h(new)} \le 123.4$$

لاحظ أن حدي التنبؤ هذين هما أضيق إلى حد ما من تلك الخاصة بتنبؤ عدد ساعات العمل لدفعة واحدة من 55 قطعة لأنها تنطوي على تنبؤ متوسط ساعات العمل لثلاث دفعات.

ونحصل على فترة التنبؤ للعدد الكلي لساعات العمل للأشواط الإنتاجية الثلاثة بضرب حدي التنبؤ لـ $\overline{Y}_{h(new)}$ بثلاثة:

$349.8 = 3(116.6) \le$ ساعات العمل الكلية $\le 3(123.4) = 370.2$

وهكذا يمكن التنبؤ بـ 90 بالمائة ثقة أن الشركة ستحتاج إلى ما يتراوح بين 350 و370 ساعة عمل لإتمام عقد يلتزم بثلاث دفعات في كل منها 55 قطعة.

(٣ ـ ٦) اعتبارات في تطبيق تحليل الانحدار

ناقشنا الآن الجزء الرئيس من استخدامات تحليل الانحدار ـ القيام باستقراءات حول معالم الانحدار، وتقدير متوسط الاستجابة لـ X معروفة، والتنبؤ بمشاهدة Y جديدة من أجل X معطاة. وبقي القليل من الملاحظات التحذيرية حول وضع تطبيقات تحليل الانحدار موضع التنفيذ.

١ ـ كثيرا ما نستخدم تحليل الانحدار للقيام باستقراءات للمستقبل، فمثلا، قد ترغب شركة وستوود تقدير ساعات العمل المتوقعة لدفعات معروفة الحجم وذلك لأغراض تخطيط الإنتاج المستقبلي. ولتطبيقات من هذا النوع فإنه من الأهمية تذكر أن مشروعية تطبيق الانحدار تعتمد على ما إذا كانت الشروط السببية الأساسية في الفترة المستقبلية ستكون مماثلة لتلك القائمة خلال الفترة التي جرى فيها تحليل الانحدار. وينطبق هذا التحذير سواء في تقدير متوسط الاستجابة أو التنبؤ بمشاهدات جديدة أو تقدير معالم انحدار.

٢ ـ عند التنبؤ بمشاهدات Y جديدة، غالبا ما نضطر إلى التنبؤ بالمتغير المستقل X نفسه. فمثلا ذكرنا سابقا التنبؤ بمبيعات شركة خلال السنة القادمة من إسقاط سكاني لعدد الأشخاص الذين يبلغون من العمر 16 فما فوق في السنة القادمة. والتنبؤ بمبيعات لشركة تحت مثل هذه الظروف هو تبنؤ مشروط، إذ يعتمد على صحة الإسقاط السكاني. ومن السهل إغفال الطبيعة الشرطية في هذا النوع من التنبؤ.

٣ ـ يعالج محذور آخر الاستقراءات المتعلقة بمستويات المتغيرالمستقل الواقعة خارج مدى المشاهدات. وللأسف كثيرا ما يحدث هذا عمليا. فالشركة التي تتنبأ بمبيعاتها عن طريق علاقة انحدارمبيعاتها على الدخل الشخصي المنتظم سنجد، في الغالب، مستوى الدخل الشخصي المنتظم المعني (مثلا للسنة القادمة) واقعا خارج مدى البيانات السابقة. وإذا لم يقع المستوى X بعيدا وراء هذا المدى فقد تتوافر ثقة مقبولة في تطبيق

تحليل الانحدار. ومن جهة أخرى، إذا وقع مستوى X بعيدا عن مدى البيانات السابقة فينبغي ممارسة الحذر الشديد لأنه لا يمكن التأكد من أن دالة الانحدار التي تصلح للبيانات السابقة ستبقى صالحة فوق مدى أوسع للمتغير المستقل.

٤ ـ الاختبار الإحصائي الذي يقود للاستنتاج أن $\beta_1 \neq 0$ لا يُرسي علاقة سبب ـ ونتيجة بين المتغيرين المستقل والتابع. فمثلا، في البيانات غير التجريبية، قد يكون المتغيران X و Y متأثرين معا بمتغيرات أخرى لم يشملها نموذج الانحدار. وهكذا فقد تُظهر بيانات عن المفردات اللغوية، عند أطفال مدرسة ابتدائية (X) وسرعة الكتابة (Y) صلة خطية واضحة. ولكن قد يكون هذا في معظمه نتيجة لعمر الطفل، وعدد سنوات الدراسة، وعوامل مشابهة تؤثر في كل من X و Y. ومن جهة أخرى فإن وجود علاقة انحدار في تجارب تحت السيطرة يشكل في الغالب دليلا جيدا لوجود علاقة سبب ـ ونتيجة.

٥ ـ وأخيرا ينبغي التنويه ثانية بظهور مشاكل خاصة عندما نرغب بتقدير متوسطات الاستجابة، أو التنبؤ بمشاهدات جديدة، من أجل عدد من المستويات المختلفة لمتغير مستقل، كما هو الحال غالبا. ومعاملات الثقة للحدين (3.32) لتقدير متوسط استجابة ولحدي التنبؤ (3.35) بمشاهدة جديدة ينطبق في حالة مستوى بمفرده لـX في عينة معطاة. وسنناقش في الفصل الخامس كيفية القيام باستقراءات متعددة من عيّنة معطاة.

(٣ ـ ٧) الحالة التي تكون فيها X عشوائية

يفترض نموذج انحدار الخطأ الطبيعي (3.1) الذي استخدم خلال هذا الفصل وسوف يستخدم باستمرار، أن قيم X ثوابت معروفة. وكنتيجة لهذا، فإن معاملات الثقة ومخاطر الأخطاء تشير إلى تكرار المعاينة مع بقاء قيم X على حالها من عينة إلى أخرى.

وكثيرا ما يكون من غير المناسب اعتبار قيم X ثوابت معروفة. خذ، مثلا، انحدار المبيعات اليومية في متجر كبير لملابس السباحة على متوسط درجة الحرارة اليومية. فمن المؤكد أنه ليس باستطاعة المتجر السيطرة على درجة الحرارة اليومية، وبالتالي ليس هناك أي معنى للتفكير في تكرار المعاينة بحيث تبقى مستويات درجة الحرارة من

عينة إلى أخرى على حالها.

في مثل هذه الحالة ربما كان من الأفضل اعتبار كل من Y و X متغيرات عشوائية. وهل يعني هذا أن جميع نتائجنا السابقة لا تنطبق هنا؟ كلا بالتأكيد. إذ يمكن تبيان أن جميع النتائج المتعلقة بالتقدير، والاختبار، والتنبؤ التي حصلنا عليها من أجل نموذج الانحدار (3.1) لا تزال تنطبق هنا إذا تحقق الشرطان التاليان:

١ـ التوزيعات الشرطية لـ Y_i علما أن X_i معطاة، هي توزيعات طبيعية ومستقلة، بمتوسط شرطي $\beta_0 + \beta_1 X_i$ وتباين شرطي σ^2.

٢ـ المتغيرات X_i هي متغيرات عشوائية مستقلة، وتوزيعها الاحتمالي $g(X_i)$ لا يتضمن المعالم β_0 ، β_1 و σ^2.

ولا يتطلب هذان الشرطان سوى أن يكون نموذج الانحدار (3.1) مناسبا لكل توزيع شرطي لـ X_i ، وأن التوزيع الاحتمالي لـ X_i لا يتضمن معالم الانحدار. فإذا تحقق هذان الشرطان فإن جميع النتائج السابقة في التقدير والاختبار والتنبؤ تبقى صحيحة بالرغم من أن المتغيرات X_i هي الآن متغيرات عشوائية. والتعديل الرئيس الذي يحدث هو في تفسير معاملات الثقة ومخاطر الخطأ ـ فعندما تكون X عشوائية، تشير هذه إلى تكرار المعاينة لقيم الأزواج (X_i, Y_i)، حيث تتغير قيم الـ X_i بالإضافة إلى قيم الـ Y_i من عينة إلى أخرى. وهكذا في المثال التوضيحي لمبيعات ملابس السباحة، يشير معامل الثقة إلى نسبة التقديرات الصحيحة بفترة إذا تكرر أخذ عينات من الحجم n للمبيعات اليومية ولدرجات الحرارة وحُسبت فترة الثقة لكل عينة ـ ويحدث تعديل آخر في قوة الاختبارات، فهي تختلف عندما يكون X متغيرا عشوائيا.

(٣ ـ ٨) أسلوب تحليل التباين في تحليل الانحدار

طورّنا حتى الآن نموذج الانحدار الأساسي وأوضحنا استخداماته الرئيسة، وعند هذه النقطة سننظر إلى العلاقة في نموذج الانحدار من منظور تحليل التباين. وهذا المنظور الجديد سوف لا يسمح لنا القيام بأي شيء جديد في نموذج الانحدار الأساسي، ولكن سيكون لأسلوب تحليل التباين خصوصيته عندما نتابع نماذج انحدار أكثر تعقيدا وأنواعا إضافية من النماذج الإحصائية الخطّية.

تجزئة مجموع المربعات الكلي

رموز أساسية. يرتكز أسلوب تحليل التباين على تجزئة مجموع المربعات ودرجات الحرية المرتبطة بمتغير الاستجابة Y. ولتوضيح الحافز لهذا الأسلوب، لنعتبر مرة أخرى مثال حجم الدفعة لشركة وستوود. إذ يوضح شكل (٣-٥)أ ساعات العمل المطلوبة لأشواط الإنتاج العشرة المعروضة سابقا في الجدول (٢-١). وتوجد اختلافات في عدد ساعات العمل، كما هو الحال في جميع البيانات الإحصائية. في الواقع لو كانت جميع المشاهدات Y_i متطابقة وعندئذٍ يكون $Y_i \equiv \overline{Y}$ فإنه لن تكون هناك أية مسائل إحصائية. واصطلاحا، يُقاس التغير في Y_i بدلالة الانحرافات:

$$Y_i - \overline{Y} \tag{3.41}$$

هذه الانحرافات مبينة في الشكل (٣-٥)أ وأحدها يحمل عنوانا ظاهرا. ومقياس التغير الكلي، ويرمز له بـ $SSTO$، هو مجموع مربعات الانحرافات (3.41):

$$SSTO = \sum (Y_i - \overline{Y})^2 \tag{3.42}$$

ويرمز $SSTO$ هنا إلى مجموع المربعات الكلي. وإذا كان $SSTO = 0$ ، فإن جميع المشاهدات متساوية. وكلما كبر $SSTO$ كلما ازداد التغير بين المشاهدات Y.

وعندما استخدمنا أسلوب الانحدار كان التغير الذي يعكس الريبة في البيانات هو ذلك المتعلق بتغير المشاهدات Y حول خط الانحدار التوفيقي:

$$Y_i - \hat{Y}_i \tag{3.43}$$

هذه الانحرافات مبينة في الشكل (٣-٥)ب. ومقياس التغير في البيانات في حالة نموذج الانحدار هو مجموع مربعات الانحرافات (3.43)، وهو الـ SSE المألوف في (2.21):

$$SSE = \sum (Y_i - \hat{Y}_i)^2 \tag{3.44}$$

ومرة أخرى، يشير SSE إلى مجموع مربعات الأخطاء. إذا كان $SSE = 0$ فإن جميع المشاهدات تقع على خط الانحدار التوفيقي وكلما كان SSE أكبر تعاظم تغير المشاهدات Y حول خط الانحدار التوفيقي.

وفي مثال شركة وستوود، نعلم من عمل سابق (جدول (٣-١)) أن:

$$SSTO = 13,660$$
$$SSE = 60$$

ما هو تعليل الفرق الكبير بين مجموعي المربعات هذين؟ والفرق، كما سنوضح بعد قليل، هو مجموع مربعات آخر:

شكل (٣-٥) تجزئة الانحرافات الكلية $Y_i - \overline{Y}$ (لم تُرسم قيم Y وفقا لسلم قياس)

(أ) (ب) (جـ) (د)

$$SSR = \sum(\hat{Y}_i - \overline{Y})^2 \tag{3.45}$$

حيث ترمز SSR إلى مجموع مربعات الانحدار. لاحظ أن SSR مجموع مربعات انحرافات، والانحرافات هي :

$$\hat{Y}_i - \overline{Y} \tag{3.46}$$

هذه الانحرافات مبينة في الشكل (٣-٥)جـ. وكل انحراف هو ببساطة الفرق بين القيمة التوفيقية على خط الانحدار ومتوسط القيم التوفيقية $\overline{Y}$. [تذكّر من (2.18) أن متوسط القيم التوفيقية $\hat{Y}_i$ هو $\overline{Y}$] وإذا كان خط الانحدار أفقيا بحيث يكون $Y_i - \hat{Y} = 0$ فإن $SSR = 0$. وفيما عدا ذلك يكون SSR موجبا.

ويمكن النظر إلى SSR كمقياس لمتغيرية الـ Y_i المتصلة بخط الانحدار، وكلما كبر SSR نسبة إلى $SSTO$، كان تأثير علاقة الانحدار في تفسير التغيرالكلي في المشاهدات Y_i أكبر.

ولدينا في مثال حجم الدفعة لشركة وستوود :

$$SSR = SSTO - SSE = 13{,}660 - 60 = 13{,}600$$

والتي توضح أن معظم التغيرالكلي في ساعات العمل يجد تفسيرا له في العلاقة بين حجم الدفعة وبين ساعات العمل.

تطوير رسمي للتجزئة. لنعتبر الانحراف الكلي $Y_i - \overline{Y}$ الكمية الأساسية التي تقيس التغير الكلي للمشاهدات Y_i. فيمكننا تفكيك هذا الانحراف كالتالي:

$$Y_i - \overline{Y} = \hat{Y}_i - \overline{Y} + Y_i - \hat{Y}_i \tag{3.47}$$

انحراف كلي	انحراف قيمة الانحدار التوفيقية حول المتوسط	انحراف حول خط الانحدار التوفيقي

وهكذا يمكن النظر إلى الانحراف الكلي $Y_i - \overline{Y}$ كمجموع مركبتين:

١- انحراف القيمة $\hat{Y}_i$ حول المتوسط $\overline{Y}$.

٢- انحراف Y_i حول خط الانحدارالتوفيقي.

ويبين الشكل (٣-٥) د هذا التفكيك لأحد المشاهدات.

والخاصة الجديرة بالملاحظة هي أن مجموع مربعات هذه الانحرافات يحقق العلاقة نفسها أي :

$$\sum(Y_i-\overline{Y})^2=\sum(\hat{Y}_i-\overline{Y})^2+\sum(Y_i-\hat{Y}_i)^2 \qquad (3.48)$$

أو، باستخدام الرموز في (3.44)، (3.42) و(3.45) نجد:

$$SSTO=SSR+SSE \qquad (3.48a)$$

ولإثبات هذه النتيجة الأساسية في تحليل التباين، نمضي كالتالي:

$$\sum(Y_i-\overline{Y})^2=\sum\left[(\hat{Y}_i-\overline{Y})+(Y_i-\hat{Y}_i)\right]^2$$

$$=\sum\left[(\hat{Y}_i-\overline{Y})^2+(Y_i-\hat{Y}_i)^2+2(\hat{Y}_i-\overline{Y})(\overline{Y}-\hat{Y}_i)\right]$$

$$=\sum(\hat{Y}_i-\overline{Y})^2+\sum(Y_i-\hat{Y}_i)^2+2\sum(\hat{Y}_i-\overline{Y})(Y_i-\hat{Y}_i)$$

والحد الأخير على اليمين يساوي الصفر، كما يمكن أن نرى لدى نشره:

$$2\sum(\hat{Y}_i-\overline{Y})(Y_i-\hat{Y}_i)=2\sum\hat{Y}_i(Y_i-\hat{Y}_i)-2\overline{Y}\sum(Y_i-\hat{Y}_i)$$

والمجموع الأول على اليمين يساوي الصفر من (2.20)، والثاني يساوي الصفر من (2.17). وهكذا نجد (3.48).

صيغ حسابية. الصيغ التعريفية لـ *SSTO*، *SSR* و *SSE* المقدمة أعلاه غير مريحة، غالبا، في الحسابات اليدوية. والصيغ التالية لـ *SSTO* و *SSR* مفيدة حسابيا ومكافئة جبريا للصيغ التعريفية:

$$SSTO=\sum Y_i^2-\frac{(\sum Y_i)^2}{n}=\sum Y_i^2-n\overline{Y}^2 \qquad (3.49)$$

$$SSR=b_1\left(\sum X_iY_i-\frac{\sum X_i\sum Y_i}{n}\right)$$

$$=b_1\left[\sum(X_i-\overline{X})(Y_i-\overline{Y})\right] \qquad (3.50a)$$

أو:

$$SSR=b_1^2\sum(X_i-\overline{X})^2 \qquad (3.50b)$$

وقد أُعطيت صيغ حسابية لـ *SSE* سابقا في (2.24).

وباستخدام نتائج مثال شركة وستوود الملخّصة في جدول (٣ـ١)، نحصل من أجل *SSR*، وباستخدام (3.50a)، على:

$$SSR = 2.0(6,800) = 13,600$$

وهذه بالطبع هي النتيجة نفسها التي حصلنا عليها سابقا بأخذ الفرق $SSTO - SSE$، باستثناء فرق طفيف، أحيانا، يعود إلى تدوير الأرقام العشرية.

تقسيم درجات الحرية

في مقابل تجزئة مجموع المربعات الكلي $SSTO$، هناك تجزئة لدرجات الحرية (اختصارا df) المتعلقة بها. لدينا $n - 1$ درجة حرية مرتبطة مع $SSTO$. فُقدت درجة حرية واحدة لأن الانحرافات $Y_i - \bar{Y}$ ليست مستقلة، فمجموعها يجب أن يساوي الصفر. وبصورة مكافئة، فقدنا درجة حرية واحدة لأن متوسط العينة $\bar{Y}$ قد استخدم كتقدير لمتوسط المجتمع.

وكما لاحظنا سابقا، يرتبط مع SSE , $n - 2$ من درجات الحرية. وفقدنا درجتي حرية لأننا قدّرنا المعلمتين β_0 و β_1 من أجل الحصول على القيم التوفيقية $\hat{Y}_i$.

ولِـ SSR درجة حرية واحدة مصاحبة له. وتوجد معلمتان في معادلة الانحدار، ولكن الانحرافات $\hat{Y}_i - \bar{Y}$ غير مستقلة لأن مجموعها يجب أن يكون صفرا، وبالتالي نخسر درجة واحدة من درجتي الحرية الممكنتين.

لاحظ أن درجات الحرية تجميعية:

$$(n - 1) = 1 + (n - 2)$$

وفي مثال شركة وستوود تكون درجات الحرية هذه:

$$9 = 1 + 8$$

متوسط المربعات

يُدعى مجموع المربعات مقسوما على درجات الحرية المرتبطة به متوسط مربعات (MS اختصارا) فمثلا يكون تباين العينة المعتاد متوسط مربعات، فهو مجموع مربعات $\sum(Y_i - \bar{Y})^2$ مقسوما على درجات الحرية المرتبطة به $n - 1$ ، ونهتم هنا بمتوسط مربعات الانحدار ويُرمز له بالرمز MSR:

$$MSR = \frac{SSR}{1} = SSR \qquad (3.51)$$

وبمتوسط مربعات الخطأ المعرَّف سابقا في (2.22) وهو:

$$MSE = \frac{SSE}{n-2} \tag{3.52}$$

وفي مثال شركة وستوود لدينا $SSR = 13,600$ و $SSE = 60$ وبالتالي:

$$MSR = \frac{13,600}{1} = 13,600$$

وكذلك حصلنا سابقا على:

$$MSE = \frac{60}{8} = 7.5$$

ملاحظة

مجموع متوسطي المربعين MSR و MSE لايساوي $SSTO \div (n\text{-}1) = 13,660 \div 9 = 1518$. وهكذا فإن متوسط المربعات غير تجميعي.

جدول تحليل التباين

الجدول الأساسي. يعرض الجدول (٣-٢) تجزئة المجموع الكلي للمربعات ودرجات الحرية المرتبطة به على شكل جدول تحليل تباين (جدول تحاين). ويبين الجدول كذلك متوسطات المربعات ذات الأهمية. وإضافة إلى ذلك، يوجد عمود لتوقعات متوسطات المربعات التي سيُستفاد منها لاحقا. وفي الجدول (٣-٣) نجد جدول التحاين لمثال شركة وستوود.

جدول معدّل. يُستخدم أحيانا، جدول تحاين يتضمن عنصر تفكيك إضافي فكما نذكر من (3.49):

$$SSTO = \sum (Y_i - \overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2$$

وفي جدول التحاين المعدّل، يُعرَّف مجموع المربعات غير المصحح الكلي ونرمز له بـ $SSTOU$ كما يلي:

$$SSTOU = \sum Y_i^2 \tag{3.53}$$

ونُعرِّف التصحيح من أجل مجموع مربعات المتوسط، ونرمز له بـ SS (تصحيح من أجل المتوسط) كما يلي:

$$SS\ (\text{تصحيح من أجل المتوسط}) = n\overline{Y}^2 \tag{3.54}$$

جدول (٣-٢) جدول تحاين لانحدارخطي بسيط.

مصدر التغير	SS	df	MS	E{MS}
انحدار	$SSR=\sum(\hat{Y}_i-\overline{Y})^2$	1	$MSR=\frac{SSR}{1}$	$\sigma^2+\beta_1^2\sum(X_i-\overline{X})^2$
خطأ	$SSE=\sum(Y_i-\hat{Y}_i)^2$	$n-2$	$MSE=\frac{SSE}{n-2}\sigma^2$	
مجموع	$SSTO=\sum(Y_i-\overline{Y})^2$	n -1		

جدول (٣ - ٣) جدول تحاين لمثال شركة وستوود.

مصدر التغير	SS	df	MS
انحدار	13,600	1	13,600
خطأ	60	8	7.5
مجموع	13,660	9	

ويبين الجدول (٣-٤) جدول التحاين المعدَّل. والهيئة العامة للجدول مقدمة في الجزء (ا) ونتائج شركة وستوود في الجزء (ب). وكلا النوعين من جداول التحاين مستخدم على نطاق واسع. وعادة سنستخدم النوع الأساسي.

توقع متوسط المربعات

كي نستطيع القيام باستقراءات تستند إلى أسلوب تحليل التباين، نحتاج معرفة القيمة المتوقعة لكل متوسط مربعات. وتخبرنا القيمة المتوقعة لمتوسط مربعات عمَّا يقدّره متوسط المربعات. وتزودنا نظرية الإحصاء بالنتائج التالية:

$$E\{MSE\}=\sigma^2 \quad (3.55)$$

$$E\{MSR\}=\sigma^2+\beta_1^2\sum(X_i-\overline{X})^2 \quad (3.56)$$

القيم المتوقعة لمتوسطات المربعات في (3.55) و(3.56) مبينة في جدول تحليل التباين في الجدول (٣-٢). لاحظ أن النتيجة (3.55) تتطابق مع عبارتنا السابقة بأن MSE مقدّر غير منحاز لـ σ^2.

جدول (٣ ـ ٤) جدول التحاين المعدل للانحدار الخطي البسيط ونتائج مثال شركة وستوود.

(أ) عام

مصدر التغير	SS	df	MS
انحدار	$SSR=\sum(\hat{Y}_i-\overline{Y})^2$	1	$MSR=\frac{SSR}{1}$
خطأ	$SSE=\Sigma(Y_i-\hat{Y}_i)^2$	n - 2	$MSE=\frac{SSE}{n-2}$
مجموع	$SSTO=\sum(Y_i-\overline{Y})^2$	n - 1	
تصحيح من أجل المتوسط	ss (تصحيح من أجل المتوسط) = $n\overline{Y}^2$	1	
مجموع غير مصحح	$SSTOU=\Sigma Y_i^2$		

(ب)مثال شركة وستوود.

مصدر التغير	SS	df	MS
انحدار	13,600	1	13,600
خطأ	60	8	7.5
مجموع	13,660	9	
تصحيح من أجل المتوسط	121,000	1	
مجموع غير مصحح	134,660	10	

وفيما يلي مضمونان مهمان لتوقعي متوسطي المربعات في (3.55) و(3.56):

١ ـ توقع MSE هو σ^2 سواء ارتبط X و Y خطيا أم لا، أي سواء أكان $\beta_1 = 0$ أم لا.

٢ ـ توقع MSR هو أيضا σ^2 عندما $\beta_1 = 0$. وعلى الوجه الآخر، عندما يكون $\beta_1 \neq 0$ يكون $E\{MSR\}$ أكبر من σ^2 لأن الحد $\beta_1^2\Sigma(X_i-\overline{X})^2$ في (3.56) لا بد أن يكون موجبا. وهكذا، فإن مقارنة MSE مع MSR تقترح نفسها لاختبار ما إذا كانت $\beta_1 = 0$ أم لا. فإذا كان MSE وMSR من الدرجة نفسها في الكبر، فإن ذلك يقترح أن $\beta_1 = 0$. وعلى الوجه الآخر، إذا كان MSR أكبر كثيرا من MSE فذلك

يقترح أن $\beta_1 \neq 0$. وهذه هي، في الحقيقة، الفكرة الأساسية التي يستند إليها اختبار تحليل التباين الذي سنناقشه فيما يلي.

ملاحظة

استنباط (3.55) يتبع من النظرية (3.11) والتي تفيد بأن $SSE/\sigma^2 \sim \chi^2(n-2)$ في نموذج الانحدار (3.1). ونجد بالتالي من الخاصية (1.39) لتوزيع مربع ـ كاي أن:

$$E\left\{\frac{SSE}{\sigma^2}\right\}=n-2$$

أو أن:

$$E\left\{\frac{SSE}{n-2}\right\}=E\{MSE\}=\sigma^2$$

ولإيجاد القيمة المتوقعة لـ MSR، نبدأ بـ (3.50b):

$$SSR=b_1^2\sum(X_i-\overline{X})^2$$

والآن، من (1.15a) لدينا:

$$\sigma^2\{b_1\}=E\{b_1^2\}-\left(E\{b_1\}\right)^2 \tag{3.57}$$

ونعلم من (3.3a) أن $E\{b_1\}=\beta_1$ ومن (3.3b) أن:

$$\sigma^2\{b_1\}=\frac{\sigma^2}{\sum(X_i-\overline{X})^2}$$

وبالتعويض في (3.57) نجد بالتالي:

$$E\{b_1^2\}=\frac{\sigma^2}{\sum(X_i-\overline{X})^2}+\beta_1^2$$

ونستنتج الآن:

$$E\{SSR\}=E\{b_1^2\}\sum(X_i-\overline{X})^2=\sigma^2+\beta_1^2\sum(X_i-\overline{X})^2$$

وأخيرا يكون $E\{MSR\}$:

$$E\{MSR\}=E\left\{\frac{SSR}{1}\right\}=\sigma^2+\beta_1^2\sum(X_i-\overline{X})^2$$

اختبار F لـ $\beta_1 = 0$ ضد $\beta_1 \neq 0$

تزودنا طريقة تحليل التباين العامة بمجموعة من الاختبارات المفيدة جدا لنماذج الانحدار (ولنماذج إحصائية خطّية أخرى). وفي حال الانحدار الخطي البسيط الذي نناقشه هنا، يزودنا تحليل الانحدار باختبار لـ:

$$\begin{aligned} H_0&: \beta_1 = 0 \\ H_a&: \beta_1 \neq 0 \end{aligned} \tag{3.58}$$

إحصاءة اختبار. يرمز F^* لإحصاءة الاختبار في طريقة تحليل التباين. وكما أوضحنا، فهو يقارن بين MSR وMSE بالطريقة التالية:

$$F^* = \frac{MSR}{MSE} \tag{3.59}$$

ويقترح الحافز المذكور سابقا والذي يستند إلى توقع متوسط المربعات في الجدول (٣-٢)، أن القيم الكبيرة لـ F^* تدعم H_a وقيم F^* القريبة من 1 تدعم H_0. وبعبارة أخرى، فإن الاختبار المناسب هو اختبار وحيد الذيل على اليمين.

توزيع F^*. كي نستطيع وضع قاعدة قرار إحصائية وفحص خواصها نحتاج معرفة توزيع المعاينة لِـ F^*. ونبدأ باعتبار توزيع المعاينة لـ F^* عندما تكون الفرضية H_0 صحيحة ($\beta_1 = 0$). وبهذا الخصوص تكون نظرية كوكران مفيدة جدا. ولأغراضنا هنا يمكن صياغة هذه النظرية كالتالي:

(3.60) إذا جاءت جميع المشاهدات Y_i وعددها n من التوزيع الطبيعي نفسه بمتوسط μ وتباين σ^2، وفككنا $SSTO$ إلى k من مجاميع المربعات SS_r، لكل منها df_r درجة حرية فعندئذٍ تكون الحدود SS_r/σ^2 متغيرات χ^2 مستقلة بـ df_r درجة حرية إذا كان:

$$\sum_{r=1}^{k} df_r = n-1$$

ونلاحظ من الجدول (٣-٢) أننا فكّكنا $SSTO$ إلى مجموعي مربعات SSR وSSE وكانت درجات الحرية لكل منهما تجميعية. وبالتالي:

إذا كان $\beta_1 = 0$، بحيث إن لكل Y_i المتوسط نفسه $\mu = \beta_0$ والتباين نفسه σ^2، فإن SSE/σ^2 و SSR/σ^2 هما متغيران χ^2 مستقلان.

لنعتبر الآن إحصاءة الاختبار F^* التي يمكن كتابتها كما يلي:

$$F^* = \frac{\frac{SSR}{\sigma^2}}{1} \div \frac{\frac{SSE}{\sigma^2}}{n-2} = \frac{MSR}{MSE}$$

ولكن لدينا عندئذٍ من نظرية كوكران وبفرض H_0 صحيحة:

$$F^* \sim \frac{\chi^2(1)}{1} \div \frac{\chi^2(n-2)}{n-2}$$

حيث المتغيران χ^2 مستقلان. وهكذا، تكون F^*، تحت H_0، نسبة متغيري χ^2 مستقلين، كل واحد منهما مقسوم على درجاته من الحرية. ولكن هذا هو تعريف المتغير العشوائي F في (1.44).

وهكذا أثبتنا أنه، تحت H_0، يتبع F^* التوزيع F، وبالتحديد التوزيع $F(1, n-2)$.

وتحت H_a يمكننا إثبات أن F^* يتبع توزيع F اللامركزي وهو توزيع معقد لا نحتاج إلى مزيد من دراسته في هذا الوقت.

ملاحظة

وحتى لو كان $\beta_1 \neq 0$ فإن SSR وSSE مستقلان و$SSE / \sigma^2 \sim \chi^2$. ولكن يتطلب كون كل من SSR/σ^2 و SSE/σ^2 متغيرا عشوائيا χ^2 أن يكون $\beta_1 = 0$.

وضع قاعدة قرار. حيث إن اختبار ذيل أيمن وF^* يتوزع تحت H_0 وفق $F(1,n-2)$، فتكون قاعدة القرار كما يلي، مع ضبط مخاطرة التورط بخطأ من النوع I عند α:

$$\begin{aligned} &\text{استنتج } H_0 \text{ إذا كان: } F^* \leq F(1-\alpha; 1, n-2) \\ &\text{استنتج } H_a \text{ إذا كان: } F^* > F(1-\alpha; 1, n-2) \end{aligned} \qquad (3.61)$$

حيث $F(1-\alpha; 1, n-2)$ هو المئين $100(1-\alpha)$ للتوزيع F المناسب.

مثال. باستخدام مثال حجم الدفعة لشركة وستوود مرة أخرى، دعنا نعيد الاختبار السابق حول β_1. وهذه المرة سنستخدم الاختبار F وبديلا القرار هما:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

وكما سبق ليكن $\alpha = 0.05$، وحيث إن $n = 10$ فإننا نحتاج إلى $F(0.95; 1,8)$ ونجد من الجدول (أ - ٤) في الملحق (أ) أن $F(0.95; 1,8) = 5.32$ وتكون قاعدة القرار:

استنتج H_0 إذا كان $F^* \leq 5.32$

استنتج H_a إذا كان $F^* > 5.32$

ولدينا من الجدول (٣-٣) أن $MSR = 13,600$ و$MSE = 7.5$. وبالتالي تكون F^*:

$$F^* = \frac{13,600}{7.5} = 1,813$$

وبما أن $F^* = 1,813 > 5.32$ فإننا نستنتج H_a أي أن $\beta_1 \neq 0$، أي توجد صلة خطّية بين ساعات العمل وحجم الدفعة. وهذه هي النتيجة نفسها التي وجدناها عند استخدام الاختبار t، وهذا الاتفاق بين النتيجتين لا بد منه وفقا لما سنراه بعد قليل.

والقيمة -P لإحصاءة الاختبار هي الاحتمال $P\{F(1,8) > F^* = 1,813\}$. ومن الجدول (أ -٤) يمكن مشاهدة أن القيمة -P أقل من 0.001 لأن $F(0.999; 1,8) = 25.4$.

تكافؤ اختبار F واختبار t. لمستوى معطى α يتكافأ جبريا اختبار F لـ $\beta_1 = 0$ ضد $\beta_1 \neq 0$ مع اختبار t ذي - الذيلين. ولرؤية هذا، لنتذكر من ($3.50b$) أن:

$$SSR = b_1^2 \sum (X_i - \overline{X})^2$$

وهكذا نستطيع كتابة:

$$F^* = \frac{SSR \div 1}{SSE \div (n-2)} = \frac{b_1^2 \sum (X_i - \overline{X})^2}{MSE}$$

وبما أن

$$S^2\{b_1\} = MSE / \Sigma (X_i - \overline{X})^2$$

فنحصل على:

$$F^* = \frac{b_1^2}{s^2\{b_1\}} = \left(\frac{b_1}{s\{b_1\}}\right)^2 \qquad (3.62)$$

والآن نعلم من مناقشة سابقة إن إحصاءة t^* لاختبار ما إذا كانت $\beta_1 = 0$ أم لا، هي من (3.17):

$$t^* = \frac{b_1}{s\{b_1\}}$$

وبالتربيع نحصل على عبارة F^* في (3.62)، وهكذا:

$$(t^*)^2 = \left(\frac{b_1}{s\{b_1\}}\right)^2 = F^*$$

وفي مسألتنا التوضيحية حسبنا لتونا $F^* = 1,813$. ومن عمل سابق لدينا $b_1 = 2.0$ و $S\{b_1\} = 0.04697$، وهكذا:

$$(t^*)^2 = \left(\frac{2.0}{.04697}\right)^2 = 1,813$$

ومقابل العلاقة بين t^* و F^* لدينا العلاقة التالية بين المئينات التي نحتاجها للتوزيعين t و F في الاختبارين $[t(1-\alpha/2;\ n-2)]^2 = F(1-\alpha; 1, n-2)$. وفي اختبارنا حول β_1 كان هذان المئينان $[t(0.975;8)]^2 = (2.306)^2 = 5.32 = F(0.95;1,8)$. تذكر أن الاختبار t هو اختبار ثنائي ـ الذيل بينما الاختبار F أحادي ـ الذيل.

وهكذا، عند مستوى معلوم α يمكننا استخدام الاختبار t أو الاختبار F لاختبار $\beta_1 = 0$ ضد $\beta_1 \neq 0$. وحينما يؤدي أحد الاختبارين إلى H_0 فإن الآخر سيؤدي إلى ذلك. والشيء نفسه بالنسبة لـ H_a. والاختبار t هو على كل حال، أكثر مرونة نظرا لإمكانية استخدامه لاختبار بدائل أحادية الجانب مثل $\beta_1(\leq \geq)\ 0$ ضد $\beta_1(> <)\ 0$ بينما لا يمكن استخدام F لمثل هذه الاختبارات.

٣ ـ ٩) طريقة اختبار خطّي عام

إن اختبار تحليل التباين لـ $\beta_1 = 0$ ضد $\beta_1 \neq 0$ هو مثال لاختبار عام لنموذج إحصائي خطّي. وسوف نوضّح طريقة الاختبار العام هذا بدلالة نموذج الانحدار الخطّي البسيط، ونقوم بذلك في هذا الوقت بسبب عمومية الطريقة ولاستخداماتنا الواسعة لها، وبسبب سهولة فهم الطريقة من خلال مسألتنا الحالية.

وتتضمن طريقة الاختبار الخطّي العام ثلاث خطوات أساسية، والتي سوف نتعرض لها الآن على التوالي.

نموذج تام

ونبدأ بالنموذج الذي يعتبر مناسبا للبيانات والذي يسمى في هذا السياق النموذج التام أو غيرالمقيد. وفي حالة نموذج الانحدار الخطّي البسيط يكون النموذج التام:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \text{نموذج تام} \tag{3.63}$$

ونوفق هذا النموذج التام باستخدام طريقة المربعات الدنيا، ونحصل على مجموع مربعات الخطأ. ومجموع مربعات الخطأ هو مجموع مربعات الانحرافات لكل مشاهدة Y_i حول تقدير لقيمتها المتوقعة. وفي هذا السياق سوف نشير إلى مجموع المربعات هذا بـ $SSE(F)$ وذلك لتوضيح أنها مجموع مربعات الخطأ للنموذج التام. ولدينا هنا:

$$SSE(F) = \sum\left[(Y_i - (b_0 + b_1 X_i)\right]^2 \tag{3.64}$$
$$= \sum (Y_i - \hat{Y}_i)^2 = SSE$$

وهكذا ففي النموذج التام (3.63) يكون مجموع مربعات الخطأ هو ببساطة SSE وهو يقيس التغيرات في المشاهدات Y_i حول خط الانحدار التوفيقي.

نموذج مخفّض

بعد ذلك نأخذ H_0 في الاعتبار. في مسألتنا هنا، لدينا:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \end{aligned} \tag{3.65}$$

ويسمى النموذج تحت الفرض بأن H_0 صحيحة، النموذج المخفض أو المقيد. عندما يكون $\beta_1 = 0$ يُختزل النموذج (3.63) إلى:

$$Y_i = \beta_0 + \varepsilon_i \qquad \text{نموذج مخفض} \tag{3.66}$$

ونوفق هذا النموذج المخفض بطريقة المربعات الدنيا، ونحصل على مجموع مربعات الخطأ لهذا النموذج المخفض، ويرمز له بـ $SSE(R)$. وعند توفيق النموذج المخفض المحدد في (3.66) يمكن تبيان أن مقدر المربعات الدنيا لـ β_0 هو $\overline{Y}$. وبالتالي

يكون تقدير القيمة المتوقعة لكل مشاهدة هو $b_0 = \bar{Y}$ ومجموع مربعات الخطأ لهذا النموذج المخفض هو:

$$SSE(R) = \sum (Y_i - b_0)^2 = \sum (Y_i - \bar{Y})^2 = SSTO \qquad (3.67)$$

إحصاءة اختبار

ومن المنطقي الآن مقارنة مجموعي مربعات الخطأ $SSE(F)$ و$SSE(R)$. ويمكن إثبات أن $SSE(F)$ لن يتجاوز أبدا $SSE(R)$:

$$SSE(F) \le SSE(R) \qquad (3.68)$$

والسبب هو أنه كلما زاد عدد المعالم في النموذج كلما أمكن توفيق النموذج للبيانات بصورة أفضل، وكلما كانت الانحرافات حول دالة الانحدار التوفيقية أقل. وإذا لم يكن $SSE(F)$ أقل كثيرا من $SSE(R)$ فإن النموذج التام لا يفسر التغير في الـ Y_i أكثر بكثير مما يفسره النموذج المخفض، وفي هذه الحالة تقترح البيانات صحة H_0 وبعبارة أخرى، إذا كان $SSE(F)$ قريبا من $SSE(R)$ فإن التغير في المشاهدات حول دالة الانحدار التوفيقية للنموذج التام يكون تقريبا مساويا للتغير حول دالة الانحدارالتوفيقية للنموذج المخفض، الأمر الذي يدعو للقول إن المعالم المضافة في النموذج التام لا تساعد، في الحقيقة، على تخفيض التغير في الـ Y_i. وهكذا فإن صغر الفرق $SSE(R)$ - $SSE(F)$ يقترح صحة H_0. ومن جهة أخرى، فإن الفرق الكبير يقترح صحة H_a، لأن المعالم الإضافية في النموذج تساعد بالفعل في تخفيض التغير في المشاهدات Y_i حول دالة الانحدار التوفيقية تخفيضا جوهريا.

وإحصاءة الاختبار الفعلية المستخدمة هي دالة في $SSE(R)$ - $SSE(F)$، وتحديدا:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \qquad (3.69)$$

وهي تتبع التوزيع F عندماتكون H_0 صحيحة. ودرجات الحرية df_R و df_F هي تلك المرتبطة مع مجموعي مربعات الخطأ للنموذجين المخفض والتام، على الترتيب. وتقود قيم F^* الكبيرة إلى H_a لأن الفرق الكبير $SSE(R)$ - $SSE(F)$ يقترح صحة H_a وتكون قاعدة القرار بالتالي:

$$\begin{aligned} &\text{استنتج } H_0 \text{ إذا كان } F^* \leq F(1-\alpha; df_R - df_F, df_F) \\ &\text{استنتج } H_a \text{ إذا كان } F^* > F(1-\alpha; df_R - df_F, df_F) \end{aligned} \tag{3.70}$$

ولاختبار ما إذا كانت $\beta_0 = 0$ أم لا، لدينا:

$$SSE(F) = SSE \qquad SSE(R) = SSTO$$
$$df_F = n - 2 \qquad df_R = n - 1$$

ونحصل عند التعويض في (3.69) على:

$$F^* = \frac{SSTO - SSE}{(n-1)-(n-2)} \div \frac{SSE}{n-2} = \frac{SSR}{1} \div \frac{SSE}{n-2} = \frac{MSR}{MSE}$$

وهي تتطابق مع إحصاءة اختبار تحليل التباين (3.59)

خلاصة

يمكن استخدام طريقة الاختبـار الخطّي العـام في اختبـارات معقـدة جـدا في النمـاذج الإحصائية الخطّية، إضافة إلى اختبارات بسيطة. والخطوات الأساسية، مرة أخرى، هي:

١- وفّق النموذج التام واحسب مجموع مربعات الخطأ $SSE(F)$

٢- وفق النموذج المخفض تحت H_0 واحسب مجموع مربعات الخطأ $SSE(R)$.

٣- استخدم إحصاءة الاختبار (3.69) وقاعدة القرار (3.70).

(٣ - ١٠) مقاييس وصفية للصلة بين X و Y في نموذج انحدار

ناقشنا الاستخدامات الرئيسة لتحليل الانحـدار ــ تقديـرا للمعـالم والمتوسـطات والتنبـؤ بمشاهدات جديدة ـ دون ذكـر "درجـة الصلـة الخطيـة" بـين X و Y، أو المصطلحـات المشابهة. والسـبب أن فـائدة التقديـرات والتنبـؤات تعتمـد علـى عـرض الفـترة وعلـى مقدارالدقة التي يحتاجها المستخدم، والتي تتغير مـن تطبيـق إلى آخـر. وبالتـالي لا يمكـن لمقياس وصفي بمفرده لـ "درجة الصلة الخطية" أن يستوعب المعلومات الأساسية المتعلقة بمسألة ما إذا كانت علاقة انحدار معطاة مفيدة في تطبيق بعينه.

ومع ذلك فهناك مناسـبات تكـون درجـات الصلـة الخطّيـة فيهـا موضـع اهتمـام لذاتها. وسنناقش الآن بإيجاز مقياسين وصفيين يُستخدمان كثيرا في التطبيقات العمليـة لوصف درجة الصلة الخطّية بين X و Y.

معامل التحديد

رأينا سابقا أن $SSTO$ يقيس التغير في المشاهدات Y_i، أو الريبة في تنبؤ Y عندما لا يؤخذ المتغير المستقل X في الحساب. وهكذا فإن $SSTO$ قياس للريبة في تنبؤ Y عندما لا نعتبر X. وبالمثل يقيس SSE التغير في Y_i عند استخدام نموذج انحدار يتضمن المتغير X. وعلى هذا فإن المقياس الطبيعي لتأثير X في تخفيض التغير في Y، أي تخفيض الريبة في تنبؤ Y هو:

$$r^2 = \frac{SSTO - SSE}{SSTO} = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \qquad (3.71)$$

ويسمى القياس r^2 معامل التحديد. وبما أن $0 \leq SSE \leq SSTO$ فنستنتج أن:

$$0 \leq r^2 \leq 1 \qquad (3.72)$$

ويمكننا تفسير r^2 كنسبة التخفيض في التغير الكلي الذي اقترن باستخدام المتغير المستقل X. وهكذا كلما كان r^2 أكبر ازداد تخفيض التغير الكلي في Y نتيجة ادخال المتغير المستقل X. وتحدث القيمتان الحدّيتان لـ r^2 كالتالي :

١ ـ إذا وقعت كل المشاهدات على خط الانحدار التوفيقي فإن $SSE = 0$ و $r^2 = 1$. وهذه الحالة معروضة في الشكل (٣-٦)أ. وهنا يُفسر المتغير المستقل X كل التغير في المشاهدات Y_i.

٢ ـ إذا كان ميل خط الانحدار التوفيقي $b_1=0$ بحيث يكون $\hat{Y}_i = \bar{Y}$ فإن $SSE=SSTO$ و $r^2 = 0$. وهذه الحالة معروضة في الشكل (٣-٦) ب. وهنا لا توجد صلة خطّية بين X و Y في بيانات العينة، ومع الانحدار الخطّي لا يساعد المتغير المستقل X في تخفيض التغير في المشاهدات Y_i.

ومن المستبعد في التطبيق العملي أن يكون r^2 مساويا 0 أو 1 ولكن بالأحرى في موضع ما بين هذين الحدين. وكلما اقترب من 1 كلما قيل عن زيادة درجة الصلة الخطّية بين X و Y.

معامل الارتباط

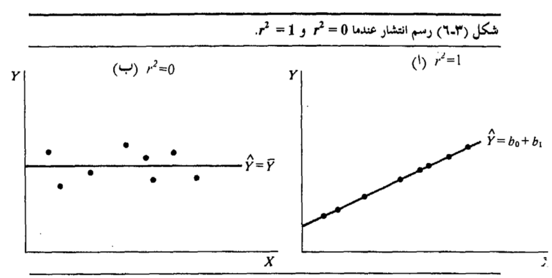
يسمى الجذر التربيعي لـ r^2 :

$$r = \pm\sqrt{r^2} \qquad (3.73)$$

معامل الارتباط. ونلحق إشارة زائد أو ناقص بهذا المقياس وفقا لما إذا كان ميـل خـط الانحدار التوفيقي موجبا أو سالبا. وهكذا يكون مدى r :

$$-1 \leq r \leq 1 \tag{3.74}$$

شكل (٣-٦) رسم انتشار عندما $r^2 = 0$ و $r^2 = 1$.



بينما يشير r^2 إلى نسبة التخفيض في متغيرية Y التي نبلغها نتيجـة لاسـتخدام معلومـات حول X، فإنه ليس للجذر التربيعي r، مثل هذا التفسير العمليـاتي الـذي لا لبْس فيـه. ومع ذلك، هناك ميل لاستخدام r بدلا من r^2 في كثير من الأعمال التطبيقية.

والجدير بالملاحظة هو أن r يمكن أن يعطي انطباعا عن علاقة "أقـرب" بـين X و Y يفوق ما يعطيه r^2. ذلك لأنه من أجل قيمـة لـ r^2، باستثناء الصفـر والواحـد، لدينا $|r| > r^2$. فمثلا $r^2 = 0.10$ تشـير إلى أن التغـيرالكلي في Y ينخفـض بنسـبة 10 بالمائة فقط عند إدخال X في الاعتبار. إلا أن $|r| = 0.32$ قد تعطي انطباعـا عـن صلـة خطّية أكبر بين X و Y.

مثال

في مثال شركة وستوود، حصلنا على $SSTO = 13{,}660$ و $SSE = 60$ وبالتالي :

$$r^2 = \frac{13{,}660 - 60}{13{,}660} = 0.996$$

وهكذا ينخفض التغير في ساعات العمل بنسبة 99.6 بالمائة عندما نأخذ حجم العينة في الاعتبار.

ومعامل الارتباط هنا هو:

$$r = +\sqrt{0.996} = +0.998$$

وأضيفت الإشارة الموجبة لأن b_1 موجبة.

صيغة حسابية لـ r

الصيغة الحسابية المباشرة لـ r والتي تزود بالإشارة المناسبة بصورة آلية هي:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2\right]^{1/2}}$$

$$= \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\left[\left(\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right)\left(\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right)\right]^{1/2}} \quad (3.75)$$

تعليقات

١ ـ تجدر ملاحظة العلاقة التالية بين b_1 و r:

$$b_1 = \left[\frac{\sum(Y_i - \bar{Y})^2}{\sum(X_i - \bar{X})^2}\right]^{1/2} r = \left(\frac{S_Y}{S_X}\right) r \quad (3.76)$$

حيث $S_Y = \left[\Sigma(Y_i - \bar{Y})^2 / (n-1)\right]^{\frac{1}{2}}$ و $S_X = \left[\Sigma(X_i - \bar{X})^2 / (n-1)\right]^{\frac{1}{2}}$ الانحرافان المعياريان للعينة من أجل المشاهدة Y و X، على الترتيب. ونلاحظ أن $b_1 = 0$ عندما $r = 0$ والعكس بالعكس. وهكذا يتضمن $r = 0$ أن خط الانحدار التوفيقي أفقي، والعكس بالعكس.

٢ ـ تنحو القيمة التي تأخذها r^2 في عينة معطاة إلى التأثر بالمسافات الفاصلة بين المشاهدات X. وهذا موجود ضمنا في (3.71). ولا تتأثر SSE بصورة منتظمة بالمسافات الفاصلة بين الـ X_i، ذلك لأنه في نموذج الانحدار (3.1) $\sigma^2\{Y_i\} = \sigma^2$ لكل مستويات X. وعلى أي حال، فكلما كانت المسافات الفاصلة بين الـ X_i في العينة

أعرض، مع كون $b_1 \neq 0$ ، اتجهت المشاهدات Y_i إلى الانتشار أكثر حول $\overline{Y}$ وبالتالي يزداد $SSTO$. ونتيجة لذلك فإنه إذ تتسع المسافات بين الـ X_i ستنحو r^2 إلى أن تكون أكبر.

٣ - يسمى مجموع مربعات الانحدار SSR غالبا "**التغير المفسَّر**" في Y. وبالتالي يسمى مجموع مربعات الرواسب SSE "**التغير غير المفسَّر**" ويسمى المجموع الكلي للمربعات "**التغير الكلي**". وعندما يُفسر المعامل r^2 بدلالة النسبة من التغير الكلي في Y التي فُسِّرت بواسطة X. وللأسف فكثيرا ما يؤخذ هذا المصطلح على وجه الحرفي، وبالتالي يُساء فهمه. ولنذكر أنه في نموذج الانحدار ليس هناك معنى ضمني يفيد أن Y تعتمد بالضرورة على X بالمعنى السببي أو التفسيري للكلمة.

٤ ـ أحيانا، تؤخذ قيمة r أو r^2 القريبة من 1 كمؤشر لإمكانية القيام باستقراءات دقيقة بما فيه الكفاية حول Y بدءا من معرفة X. وكما ذكرنا سابقا، تعتمد فائدة علاقة الانحدار على عرض فترة الثقة أو التنبؤ وحاجتنا بالذات إلى الدقة، وهي تتغير من تطبيق إلى آخر، وبالتال لا يشكل مقياس بمفرده مؤشرا مناسبا لفائدة علاقة الانحدار.

٥ - لا تتضمن نماذج الانحدار أي معلمة نقدرها بـ r أو r^2. فهذان المعاملان هما ببساطة مقياسان وصفيان لدرجة الصلة الخطية بين X و Y في مشاهدات العينة والتي قد تكون أو لا تكون مفيدة في أي ظرف بعينه.

(٣ - ١١) مدخلات ومخرجات حاسب

كان المعتاد أن تكون حسابات الانحدارشاقة ومملة، خاصة عندما يكون عدد المشاهدات كبيرا وعندما تكون هناك متغيرات مستقلة عديدة. واليوم يمكن استخدام الحاسب الآلي بسهولة لإجراء حسابات الانحدار بالاستفادة من أي حزمة من العديد من حزم البرامج المتوافرة. وكذلك يحتوي عدد من الحاسبات الشخصية على روتين انحدار.

ويختلف إدخال البيانات من برنامج إلى آخر. ففي بعضها، تدخل المشاهدات X و Y كمجموعتين منفصلتين. وفي حالات أخرى، يكون إدخال البيانات على الشكل X_1 ، Y_1 ، X_2 ، Y_2 و إلخ.

وكذلك تختلف مُخرجات الحاسب الآلي من حزمة برامج إلى أخرى. ويوضّح الشكل (٣-٧) هيئة تقليدية، للمخرجات عند توفيق نموذج انحدار خطّي بسيط لبيانات شركة وستوود في الجدول (٢-١) باستخدام حزمة الحاسب $SPSS^{X}$ (مرجع 3.1). طُبعت المشاهدات العشر لحجم الدفعة وساعات العمل في القمة، وهذا يسمح بالتحقق من أن المشاهدات أدخلت إلى الحاسب بشكل دقيق. بعد ذلك أعطيت r ، r^2 و $\sqrt{MSE}$ ويليهما معاملا الانحدار المقدّران وتقدير الانحراف المعياري لـ b_1 بالإضافة إلى إحصاءة الاختبار F^*. وأخيرا أعطيت إحصاءات وصفية للمتغيرين X و Y وتلاها جدول تحليل التباين.

وفي الشكل (٣-٧) أشرنا إلى المخرجات بدلالة الرموز المستخدمة في هذا الكتاب. وتتفق جميع النتائج في الشكل (٣-٧) مع حساباتنا السابقة، باستثناء ما كان منها في عدد الأرقام العشرية.

ولا تتضمن مخرجات حزمة الحاسب الموضّحة في الشكل (٣-٧) $s\{b_0\}$ ، الانحراف المعياري المقدَّر لـ b_0. وعلى كل حال، يمكن حساب هذا التقدير بسهولة من البيانات المعطاة في مُخرجات الحاسب. لاحظ بهذا الخصوص أن حد المقام $\sum(X_i - \bar{X})^2$ في (3.21) يساوي $(n-1)S_X^2$ وأن S_X معطى في مخرجات الحاسب.

تختلف مطبوعات مخرجات الحاسب الآلي لبرامج تحليل الانحدار اختلافا كبيرا في هيئتها من برنامج لآخر. إضافة إلى أنه قد تحدث اختلافات في النتائج المحسوبة وذلك لأن برامج الحزم المختلفة لا تسيطر على أخطاء تدوير الأرقام العشرية بالجودة نفسها.

وقبل استخدام برنامج حاسب آلي للمرة الأولى، فإنه من المستحسن التحقق باستخدامه في بيانات نتائجها الدقيقة معروفة.

شكل (٣-٧) قطعة من مُخرجات حاسب لتشغيله انحدار بيانات شركة وستورد ($SPSS^X$، مرجع[3.1]).

```
     Xi                          Yi
     |              VARIABLES   /
1 SIZE ↓            2 HOURS   ↙
   30.0000             73.0000
   20.0000             50.0000
   60.0000            128.0000
   80.0000            170.0000
   40.0000             87.0000
   50.0000            108.0000
   60.0000            135.0000
   30.0000             69.0000
   70.0000            148.0000
   60.0000            132.0000

DEPENDENT VARIABLE..    HOURS

VARIABLE(S) ENTERED ON STEP NUMBER  1..     SIZE

MULTIPLE R          0.99780 ←—r
R SQUARE            0.99561 ←—r²

STANDARD ERROR      2.73861 ←—√MSE
---------------- VARIABLES IN THE EQUATION ------------------
VARIABLE            B                       STD ERROR B        F
SIZE          2.000000 ←— b1          s{b1} —→ 0.04697    1813.333 ←—F*
(CONSTANT)    10.00000 ←— b0

VARIABLE          MEAN           STANDARD DEV        CASES
SIZE       X̄ —→ 50.0000     sX —→ 19.4365           10
HOURS      Ȳ —→ 110.0000    sY —→ 38.9587           10 ⇐ n

ANALYSIS OF VARIANCE     DF      SUM OF SQUARES           MEAN SQUARE
REGRESSION               1.   SSR —→ 13600.00000   MSR —→ 13600.00000
RESIDUAL ←—Error         8.   SSE —→ 60.00000      MSE —→ 7.50000
```

مراجع وردت في النص

[3.1] SPSS^X *User's Guide*, 2nd ed. Chicago : SPSS, Inc., 1986.

مسائل

(٣-١) طالب، يعمل في دورة تدريب صيفية في مكتب بحوث اقتصادية لشركة كبيرة، يدرس العلاقة بين مبيعات منتج (Y، بملايين الدولارات) وعدد السكان

(X.بملايين الأشخاص) في مناطق تسويق الشركة الـ 50. وقد استخدم نموذج انحدار الخطأ الطبيعي (3.1). ويرغب الطالب أولا اختبار ما إذا كانت هناك صلة خطّية بين Y وX. واستعان الطالب ببرنامج انحدار خطّي بسيط وحصل على المعلومات التالية عن معاملات الانحدار:

المعلمة	القيمة المقدّرة	95 بالمائة حدي ثقة	
الجزء المقطوع	7.43119	-1.18518	16.0476
الميل	0.755048	0.452886	1.05721

ا - استنتج الطالب من هذه النتائج أنه توجد رابطة خطّية بين Y و X. هل الاستنتاج مبرر؟ ما هو مستوى المعنوية المتضمن هنا؟

ب ـ شكك أحدهم بالإشارة السالبة لحد الثقة الأدنى للجزء المقطوع، مشيرا إلى أن المبيعات بالدولار لايمكن أن تكون سالبة حتى لو كان عدد السكان في المقاطعة صفرا. ناقش.

(٣-٢) في اختبار للبدائل $H_0: \beta_1 \leq 0$ ضد $H_a: \beta_1 > 0$ استنتج محلل الفرضية H_0. هل يتضمن القرار أنه لاتوجد صلة خطّية بين X و Y؟ اشرح.

(٣-٣) استلم عضو في فريق طلاّبي يلعب لعبة تسويق نشطة المخرجات التالية لحاسب آلي وذلك عند دراسة العلاقة بين تكاليف الإعلان (X) والمبيعات (Y) لأحد منتجات الفريق:

معادلة الانحدار المقدَّرة: $\hat{Y} = 350.7-0.18X$

القيمة -P ثنائية الجانب للميل المقدر: 0.91

صرح الطالب: "الرسالة التي حصلت عليها هنا هي أنه كلما زدنا الإنفاق على الدعاية لهذا المنتج، بعنا وحدات **أقل**! "علّق.

(٣-٤) بالعودة إلى مسألة **المعدل التراكمي** رقم (٢-١٧)، بعض النتائج الإضافية هي:

$b_0 = -1.700$ ، $S\{b_0\} = 0.7267$ ، $b_1 = 0.8399$ ، $S\{b_1\} = 0.144$ ، $MSE = 0.1892$

ا ـ أوجد 99 بالمائة فترة ثقة لـ β_1. فسّر فترة ثقتك. هـل تتضمـن الصفـر؟ لماذا يمكن أن يهتم مدير القبول فيما إذا كانت فترة الثقة تتضمن الصفر؟

ب ـ اختبر، باستخدام إحصاءة الاختبار t^*، ما إذا كانت توجد رابطة خطّية أم لا بين درجـة الطـالب في اختبـار الدخـول X ومعدلـه الـتراكمي في نهاية السنة الأولىY. استخدام مستوى معنويــة 0.10 واكتـب البديلـين وقاعدة القرار والنتيجة.

جـ ـ ما هي القيمة -P لاختبارك في الجـزء (ب)؟ كيـف تدعـم هـذه القيمـة النتيجة التي وصلت اليها في الجزء (ب)؟

(٣–٥) بالعودة إلى مسألة **صيانة الحاسبات اليدوية** (٢–١٨). بعض النتائج الإضافيـة هـي b_0=-2.3221 ، $s\{b_0\}$=2.564 ، b_1=14.738 ، $s\{b_1\}$=0.519 و MSE = 20.086.

ا ـ قدّر التغير في متوسـط زمـن الخدمـة عندمـا يـزداد عـدد المكـائن المصانـة واحدا. استخدم 90 بالمائة فترة ثقة. فسّر فترة ثقتك.

ب ـ قم باختبار t لتحديد ما إذا كانت توجد رابطـة خطّيـة أم لا بـين X و Y هنا؛ اضبط المخاطرة α عند مستوى 0.10، أعرض البديلين وقاعدة القرار والنتيجة. ماهي القيمة -P لاختبارك؟

جـ ـ هل النتائج في الجزئين (أ) و (ب) متسقة؟ وضّح.

د ـ اقترح المصنع أن لايزداد متوسط الوقت المطلوب بأكثر من 14 دقيقة لكل آلـة إضافية تتم صيانتها خلال نداء خدمة. قـم باختبـار لتقريـر مـا إذا كـان هـذ المعيار محققا في تراي ـ ستي. اضبط مخاطرة الخطأ من النوع الأول عنـد 0.05 اذكر البديلين وقاعدة القرار والنتيجة. ماهي القيمة -P للاختبار؟.

هـ ـ هل تعطي b_0 هنا أية معلومات مناسبة عن وقت البداية الفعليـة للصيانـة أي عن الوقت المطلوب الذي يفصل بين وصـول فـي الصيانـة إلى موقـ الآلات وبين بداية عمل الصيانة الفعلي؟

(٣–٦) بالعودة إلى مسألة تكسر **الشحنات الجوية** (٢–١٧).

ا ـ قدّر β_1 بـ 95 بالمائة فترة ثقة. فسّر تقديرك بفترة.

ب ـ قم باختبار t لتقرير ما إذا كانت توجد رابطة خطّية أم لا بين عدد المرات التي يحول فيها الصندوق من طائرة إلى أخرى (X) وعدد الأنبولات المكسّرة (Y). استخدم مستوى معنوية 0.05 أعرض البديلين وقاعدة القرار والنتيجة. ما هي القيمة -P للاختبار؟

جـ ـ تمثل β_0 هنا متوسط عدد الأنبولات المكسرة عندما لايحدث أي تحويلات للشحنة أي عندما $X = 0$. أوجد 95 بالمائة فترة ثقة لـ β_0 وفسرها.

د ـ اقترح مستشار، معتمدا على خبرة سابقة، أنه ينبغي ألا يتجاوز متوسط عدد الأنبولات المكسّرة 9 عندما لاتتعرض الشحنة لتحويل. قم باختبار مناسب مستخدما $\alpha = 0.025$. أذكر البديلين و قاعدة القرار والنتيجة. ماهي القيمة -P للاختبار؟

هـ ـ أوجد قوة اختبارك في الجزء (ب) إذا كانت β_1=2.0 فعلا. إفترض $\sigma\{b_1\} = 0.50$. أوجد كذلك قوة اختبارك في الفقرة (د) إذا كانت $\beta_0 = 11$. افترض أن $\sigma\{b_0\} = 0.75$ فعلا.

(٣-٧) بالعودة إلى مسألة **صلابة البلاستيك** (٢-٢٠).

ا ـ قدِّر التغير في متوسط الصلابة عندما يزداد الوقت المنصرم بساعة واحدة. استخدم 99 بالمائة فترة ثقة، فسّر تقديرك بفترة.

ب ـ صرح منتج البلاستيك أنه ينبغي أن يزداد متوسط الصلابة بوحدتي برينل لكل ساعة. قم باختبار ثنائي ـ الجانب لتقرير ما إذا كانت هذه المواصفة محققة. استخدم $\alpha = 0.01$. اذكر البدائل وقاعدة القرار والنتيجة ما هي القيمة ـ P للاختبار؟

جـ- أوجد قوة اختبارك في الجزء (ب) إذا اجتيز المعيار الحقيقي بمقدار 0.3 وحدة برينل لكل ساعة. افترض أن $\sigma\{b_1\}$ =0.1.

(٣-٨) بالعودة إلى شكل (٣-٧) لمثال شركة وستوود. نصَح مستشار أن زيادة وحدة واحدة في حجم الدفعة ينبغي أن يتطلب زيادة 1.8 في العدد المتوقع لساعات العمل للمفردة المنتجة المعطاة.

ا ـ قم باختبار لتقرير ما إذا كانت الزيادة في العدد المتوقع لساعات العمل في شركة وستوود تتفق وهذا المعيار أم لا. استخدم α =0.05، اذكر البديلين وقاعدة القرار والنتيجة.

ب ـ أوجد قوة اختبارك في الجزء (أ) إذا كان معيار المستشار قد جرى تجاوزه فعلا بـ 0.1 ساعة. إفرض $\sigma\{b_1\}$ =0.05.

جـ ـ لماذا لاتناسب $F^* = 1813.333$، المبينة في مطبوعة الحاسب الاختبار في الجزء (ا)؟

(٣-٩) بالعودة إلى شكل (٣-٧). سأل طالب، ملاحظا أن $s\{b_1\}$ قد أُعطي في مطبوعة الحاسب، لماذا لم يُعط أيضا $s\{\hat{Y}_h\}$. ناقش.

(٣-١٠) لكل من الأسئلة التالية، وضّح ما إذا كانت فترة ثقة لمتوسط استجابة أم فترة تنبؤ لمشاهدة جديدة هي الأنسب.

ا ـ ماذا سيكون مستوى الرطوبة في البيت الزجاجي غدا عندما نضع مستوى درجة الحرارة عند 31 °C؟.

ب ـ ما هو متوسط مصروف العائلات، التي دخلها المنتظم 23,500 $، على الطعام خارج المنزل؟

جـ ـ كم عدد الكيلو واط ـ ساعة من الكهرباء التي ستستهلك الشهر القادم من قبل المستهلكين التجاريين والصناعيين في منطقة خدمات توين ستيز، علما أن الرقم القياسي لنشاط الأعمال التجارية والصناعية في المنطقة يبقى على مستواه الحالي؟

(٣-١١) يسأل شخص عما إذا كان يوجد فرق بين "متوسط الاستجابة عند $X = X_h$" و"متوسط m مشاهدة جديدة عند $X = X_h$". أجب.

(٣-١٢) هل يمكن تحقيق اقتراب $\sigma^2\{Y_{h(new)}\}$ في (3.36) من الصفر كلما أصبحت كبيرة؟ هل هذه هي الحال أيضا بالنسبة لـ $\sigma^2\{\hat{Y}_h\}$ في (3.28b) ؟ ما هو المعنى الذي يتضمنه هذا الفرق.

(٣-١٣) بالعودة إلى مسألتي **المعدل التراكمي** (٢-١٧) و(٣-٤)

ا ـ أوجد 95 بالمائة فترة تقدير لمتوسط المعدل التراكمي لطلاّب السنة الأولى الذين حصلوا في اختبار الدخول على درجة 4.7. فسّر فترة ثقتك.

ب ـ حصلت ماري جون على علامة 4.7 في اختبار الدخول. تنبأ بمعدلها التراكمي في السنة الأولى مستخدما 95 بالمائة فترة تنبؤ. فسّر نتائجك.

جـ ـ هل فترة التنبؤ في الجزء (ب) أوسع من فترة الثقة في الجزء (أ) ؟ هل ينبغي ذلك؟

(٣-١٤) بالعودة إلى مسألتي **صيانة الحاسبات** (٢-١٨) و(٣-٥).

ا ـ أوجد 90 بالمائة فترة تقدير لمتوسط زمن الخدمة للنداءات التي تم فيها خدمة ست آلات. فسّر فترة ثقتك.

ب ـ أوجد 90 بالمائة فترة تنبؤ لزمن الخدمة لنداء قادم تجري فيه صيانة ست آلات؟ هل فترة تنبؤك أوسع من فترة الثقة المقابلة في الجزء (ا)؟ هل ينبغي ذلك؟

جـ ـ افترض أن الإدارة ترغب في تقدير زمن الخدمة المتوقع للآلة الواحدة في حالة نداءات تجري فيها صيانة ست آلات. أوجد فترة ثقة مناسبة بتحويل الفترة التي حصلت عليها في الجزء (ا). فسّر فترة الثقة المحوَّلة.

(٣-١٥) بالعودة إلى مسألة **تكسر الشحنات** الجوية (٢-١٩). و

ا ـ بسبب تغير مسارات الطائرات، ربما يتكرر تحويل الشحنات أكثر من السابق. قدّر متوسط عدد الكسور للأعداد التالية من التحويلات $X = 2 , 4$. استخدم 99 بالمائة فترتي ثقة منفصلتين. فسّر نتائجك.

ب ـ تتضمن الشحنة القادمة تحويلين. أوجد 99 بالمائة فترة تنبؤ لعدد الأنبولات المكسّرة لهذه الشحنة. فسّر فترة تنبؤك.

جـ ـ في الأيام العديدة القادمة ستُرسل ثلاث شحنات بصورة مستقلة. كل شحنة تتطلب تحويلين. أوجد 99 بالمائة فترة ثقة لمتوسط عدد

الأنبولات المكسّرة في الشحنات الثلاث. حول هذه الفترة إلى 99 بالمائة فترة تنبؤ للعدد الكلي للأنبولات المكسّرة في الشحنات الثلاث.

(٣-١٦) بالعودة إلى مسألة **تصلب البلاستيك** (٢-٢٠).

ا - أوجد 98 بالمائة فترة ثقة لمتوسط صلابة الوحدات المشكّلة بزمن منصرم مدته 30 ساعة. فسّر فترة الثقة.

ب - أوجد 98 بالمائة فترة تنبؤ لصلابة وحدة جديدة بزمن منصرم مدته 30 ساعة.

جـ - أوجد 98 بالمائة فترة تنبؤ لمتوسط صلابة 10 وحدات اختبار مشكّلة حديثا كل واحدة منها بزمن منصرم مدته 30 ساعة.

د - هل فترة التنبؤ في (جـ) أضيق من تلك في (ب) ؟ هل ينبغي ذلك؟

(٣-١٧) قام محلل بتوفيق نموذج انحدار الخطأ الطبيعي (3.1)، وأجرى اختبار F لـ $\beta_1 = 0$ ضد $\beta_1 \neq 0$ وكانت القيمة-P للاختبار 0.033، وقد استنتج المحلل $H_a : \beta_1 \neq 0$. هل كان المستوى α المستخدم من قبل المحلل أكبر أو أصغر من 0.033؟ ولو كانت قيمة المستوى α هي 0.01. فماذا ينبغي أن تكون النتيجة المناسبة؟

(٣-١٨) لإجراء اختبارات إحصائية حول المعلمة β_1 ، لماذا يكون اختبار t أكثر شيوعا من الاختبار F؟

(٣-١٩) عند اختبار ما إذا كان $\beta_1 = 0$ أم لا، لماذا يكون الاختبار F اختبارا أحادي - الجانب مع أن H_a تتضمن كلا من $\beta_1 < 0$ و $\beta_1 > 0$ ؟ [ارشاد: ارجع إلى (3.56)].

(٣-٢٠) يسأل طالب فيما إذا كان r^2 مقدّرا خطّيا لأي معلمة في نموذج انحدار الخطأ الطبيعي (3.1). أجب.

(٣-٢١) في بعض الأحيان تُفسر قيمة لـ r^2 قريبة من الواحد على أنها تتضمن أن العلاقة بين Y و X قريبة قربا كافيا بحيث يمكن القيام بتنبؤ دقيق عن Y بدءا من معرفة X. هل يُعتبر هذا التضمن نتيجة لتعريف r^2 لامناص منها؟

(٣-٢٢) باستخدام نموذج انحدار الخطأ الطبيعي (3.1) في تجربة أمان هندسية، وجد باحث في المشاهدات العشر الأولى أن r^2 كان صفرا. هل يمكن أن يكون r^2 غير الصفر للمجموعة الكاملة من 30 مشاهدة؟ هل يمكن لـ r^2 أن لايكون صفرا للمشاهدات العشر الأولى، ويكون مع ذلك صفرا للمشاهدات الثلاثين كافة؟ إشرح.

(٣-٢٣) بالعودة إلى مسألتي **المعدل التراكمي** (٢-١٧) و(٣-٤). إليك بعض النتائج الحسابية الإضافية $SSE = 3.406$ و $SSR = 6.434$.

ا ـ اكتب جدول التحاين.

ب ـ ماالذي يقدّره MSR في جدول تحاينك ؟ ما الذي يقدّره MSE؟ تحت أية شروط تقدّر MSR و MSE الكمية نفسها؟

جـ ـ قم باختبار F لما إذا كان $\beta_1 = 0$ أم لا، اضبط المخاطرة α عند 0.01 اذكر البديلين وقاعدة القرار والنتيجة.

د ـ ما هي القيمة المطلقة للتخفيض في تغير Y عند إدخال X في نموذج الانحدار؟ ما هو التخفيض النسبي؟ ما هو اسم القياس الأخير.

هـ ـ أوجد r وألحق بها الإشارة المناسبة.

و ـ أي المقياسين r^2 أم r يتمتع بتفسير عملياتي أكثر وضوحا ؟ اشرح.

(٣-٢٤) بالعودة إلى مسألتي **صيانة الحاسبات اليدوية** (٢- ١٨) و(٣-٥) إليك بعض النتائج الحسابية الإضافية $SSE = 321.396$ و $SSR = 16{,}182.604$.

ا ـ ضع جدول تحاين أساسي في هيئة الجدول (٣-٢). أي العناصر في جدولك تجميعي؟ ضع أيضا جدول تحاين في هيئة الجدول (٣-٤)أ. كيف يختلف الجدولان؟

ب ـ قم باختبار F لتحديد ما إذا كانت توجد رابطة خطّية أم لا بين الوقت المستغرق وعدد المكائن المصانة، استخدم $\alpha = 0.10$. اذكر البديلين وقاعدة القرار والنتيجة.

جـ ـ وبصورة نسبية، كم ينخفض التغير الكلي في عـدد الدقـائق المبذولـة في نداء خدمـة عنـد إدخـال عـدد الآلات المصانـة في التحليـل؟ هـل هـذا التخفيض صغير أم كبير نسبيا؟ ما اسم هذا المقياس ؟

د ـ احسب r وألحق بها الإشارة المناسبة.

هـ ـ أي القياسين r أم r^2 له تفسير عملياتي أكثر وضوحا؟

(٣ـ٢٥) بالعودة إلى مسألة **تكسر الشحنات** الجوية (٢ـ١٩).

ا ـ ضع جدول تحاين. أي العناصر تجميعي ؟

ب ـ قم باختبار F لتقرر ما إذا كانت توجد رابطـة خطّيـة أم لا بـين عـدد مرات تحويل الصندوق وعدد الأنبولات المكسّرة، اضبط α عند 0.05، اذكر البدائل وقاعدة القرار والنتيجة.

جـ ـ أوجد إحصاءة الاختبار t^* للاختبار في الجزء (ب) وأثبـت عدديـا أنهـا مكافئة للإحصاءة F^* في الجزء (ب).

د ـ احسب r و r^2. ما هي نسـبة التغير في Y الـذي يفسّـرها إدخـال X في نموذج الانحدار؟

(٣ـ٢٦) بالعودة إلى مسألة **تصلب البلاستيك**.

ا ـ اكتب جدول التحاين.

ب ـ اختبر مستخدما الاختبار F، ما إذا كانت توجد رابطة خطّية أم لا بـين تصلب البلاستيك والمدّة المنصرمة. استخدم $\alpha = 0.01$ اذكـر البديلـين وقاعدة القرار والنتيجة.

جـ ـ ارسم الانحرافات $Y_i - \hat{Y}_i$ مقابل X_i في رسم بيـاني. وارسـم الانحرافـات $\hat{Y}_i - \bar{Y}$ في مقابل X_i في رسم بياني آخر. من رسميك البيـانيين، هـل تبدو SSR أم SSE المركّبة الأكبر في $SSTO$؟

د ـ أحسب r و r^2.

(٣ـ٢٧) بالعودة إلى مسألة **كتلة العضلة** (٢ـ٢٥).

ا ـ قم باختبار لتقرير ما إذا كانت توجد رابطة خطّية سلبية أم لا بين مقدار كتلة العضلة والعمر، اضبط مخاطرة الخطأ من النوع الأول عند 0.05. اعرض البديلين، وقاعدة القرار والنتيجة. ما هي القيمة -P للاختبار ؟

ب ـ القيمة -P ثنائية الجانب لاختبار ما إذا كانت $\beta_1 = 0$ هي +0. هل يمكنك الآن استنتاج أن b_0 تقدم معلومات مناسبة عن مقدار كتلة العضلة لطفلة عند الولادة؟

جـ ـ قدّر بـ 95 بالمائة فترة ثقة، الفرق في كتلة العضلة المتوقعة لنساء تختلف أعمارهن بسنة واحدة. لماذا تكون معرفة الأعمار بالتحديد غير ضرورية للقيام بهذا التقدير؟

(٣ـ٢٨) بالعودة إلى مسألة **كتلة العضلة** (٢ـ٢٥).

ا ـ أوجد 95% فترة ثقة لمتوسط كتلة العضلة لنساء بلغن الستين من العمر. فسّر فترة ثقتك.

ب ـ أوجد 95% فترة ثقة لمتوسط كتلة العضلة لامرأة عمرها 60 عاما. هل فترة التنبؤ دقيقة نسبيا؟

(٣ـ٢٩) ارجع إلى مسألة **كتلة العضلة** (٢ـ٢٥).

ا ـ ارسم الانحرافات $Y_i - \hat{Y}_i$ مقابل X_i في رسم بياني. وارسم الانحرافات $\hat{Y}_i - \bar{Y}$ في مقابل X_i في رسم بياني آخر، من رسميك البيانيين هل تبدو SSE أم SSR المركّبة الأكبر في $SSTO$؟

ب ـ اكتب جدول التحاين.

جـ ـ إختبر ما إذا كان $\beta_1 = 0$ أم لا مستخدما الاختبار F مع $\alpha = 0.10$. اذكر البديلين، وقاعدة القرار والنتيجة.

د ـ ما هي النسبة من التغير الكلي في كتلة العضلة التي تبقى غير مفسّرة عند إدخال العمر في التحليل؟ هل هذه النسبة صغيرة أم كبيرة نسبيا؟

هـ ـ أوجد r^2 و r.

(٣-٣٠) بالعودة إلى مسألة **معدل السرقة** (٢-٢٦).

ا - اختبر ما إذا كانت توجد رابطة خطّية أم لا بين معدّل السرقة وكثافة السكان مستخدما الاختبار t مع $\alpha = 0.01$. اذكر البديلين وقاعدة القرار والنتيجة. ما هي القيمة -P للاختبار؟

ب - اختبر ما إذا كانت $\beta_0 = 0$ أم لا؛ اضبط مخاطرة الخطأ من النوع I عند 0.01. اذكر البدائل وقاعدة القرار والنتيجة. لماذا يمكن الاهتمام باختبار ما إذا كانت $\beta_0 = 0$ أم لا ؟

جـ - قدّر β_1 بـ 99 بالمائة فترة ثقة. فسّر فترة ثقتك.

(٣-٣١) بالعودة إلى مسألة **معدل السرقة** (٢-٢٦).

ا - ضع جدول التحاين.

ب - نفذ الاختبار في المسألة (٣-٣٠)أ مستخدما الاختبار F وبين التكافؤ العددي لإحصائي الاختبار، ولقاعدتي القرار. هل القيمة -P في الاختبار F هي نفسها في الاختبار t؟

جـ - كم ينخفض التغير الكلي لمعدل السرقة عند إدخال كثافة السكان في التحليل؟ هل التخفيض كثير أم قليل نسبيا ؟

د - أوجد r.

(٣-٣٢) بالعودة إلى مسألتي **معدل السرقة** (٢-٢٦) و(٣-٣٠)، افترض أننا نريد تنفيذ الاختبار (٣-٣٠)أ باستخدام اختبار خطّي عام.

ا - اذكر النموذجين التام والمخفض.

ب - أوجد (١) $SSE(F)$ ، (٢) $SSE(R)$، (٣) df_F، (٤) df_R، (٥) إحصاءة الاختبار F^* للاختبار الخطي العام و (٦) قاعدة القرار.

جـ - هل إحصاءة الاختبار F^* وقاعدة القرار للاختبار الخطي العام مكافئتان عدديا لتلك في مسألة (٣-٣٠)أ؟

(٣-٣٣) عند تطوير دالة تكلفة تجريبية من بيانات لوحظت في تجربة كيميائية معقدة، استخدم محلل نموذج انحدار الخطأ الطبيعي (3.1). وقد فُسرت β_0 هنا، كتكلفة إعداد التجربة. افترض المحلل أن هذه التكلفة ينبغي أن تكون 7500 $ ويرغب اختبار الفرضية بواسطة اختبار خطّي عام.

ا ـ أذكر القرارين البديلين لهذا الاختبار.

ب ـ حدّد النموذجين التام والناقص.

جـ ـ بدون أية معلومات إضافية، هل يمكنك القول ماذا ستساوي في اختبار المحلل، الكمية $df_R - df_F$ في إحصاءة الاختبار (3.69)؟ إشرح.

(٣-٣٤) بالعودة إلى مسألة **المعدل التراكمي** (٢-١٧).

ا ـ أيها أكثر منطقية هنا أن نعتبر الـ X_i ثوابت معروفة أم متغيرات عشوائية؟ اشرح.

ب ـ إذا اعتبرت الـ X_i متغيرات عشوائية فهل لهذا أي تأثير على فترات التنبؤ لمتقدمين جدد ؟ إشرح.

(٣-٣٥) بالعودة إلى مسألتي **صيانة الحاسبات اليدوية** (٢-١٨) و(٣-٥). كيف سيختلف معنى معامل الثقة في المسألة (٣-٥)أ إذا اعتُبر المتغير المستقل متغيرا عشوائيا وبقيت الشروط في (3.40) قابلة للتطبيق؟

تمارين

(٣-٣٦) اشتق الخاصية في (3.6) للمقادير K_i.

(٣-٣٧) بين أن b_0 كما عُرِّفت في (3.19) مقدّر غير منحاز لـ β_0.

(٣-٣٨) استنبط العبارة في (3.20b) الخاصة بتباين b_0 مستخدما نظرية (3.30). وضّح كذلك كيف أن التباين (3.20b) هو حالة خاصة من التباين (3.28b).

(٣-٣٩) (يُحتاج للتفاضل).

ا ـ أوجد دالة الإمكانية لمشاهدات العينة $Y_1, \ldots, Y_n$ علما أن $X_1, \ldots, X_n$ معطاة، وذلك عند تحقق الشروط (3.40).

ب ـ أوجد مقدّرات الإمكانية العظمى لـ β_0 ، β_1 و σ^2. هل يبقى مقدّرا β_0 و β_1 كما هما في (2.27) عندما تكون الـ X_i مثبتة.

(٣ـ٤٠) افترض أن نموذج انحدار الخطأ الطبيعي (3.1) قابل للتطبيق باستثناء أن تباين الخطأ غير ثابت؛ وإنما يزداد مع تزايد X. هل لايزال $\beta_1 = 0$ يتضمن عدم وجود رابطة خطّية بين X و Y؟ عدم وجود رابطة بين X و Y؟ وضّح.

(٣ـ٤١) استنبط العبارة الخاصة بـ SSR في (3.50b).

(٣ـ٤٢) في دراسة انحدار على نطاق ضيق، حصلنا على خمس مشاهدات Y في مقابل $X = 1, 4, 10, 11, 14$. افترض أن $\sigma = 0.6$ ، $\beta_0 = 5$ و $\beta_1 = 3$

ا ـ ما هما القيمتان المتوقعتان لـ MSR و MSE هنا؟

ب ـ لأغراض تحديد ما إذا كانت توجد علاقة انحدار أم لا، هل كان سيكون من الأفضل أم الأسوأ قياس المشاهدات الخمس عند $X = 6,$ $7, 8, 9, 10$؟ لماذا؟

هل يبقى الجواب نفسه إذا كان الغرض الرئيس هو تقدير متوسط الاستجابة عند $X = 8$ ؟ ناقش.

(٣ـ٤٣) افترض أن نموذج انحدار الخطأ الطبيعي (3.1) قابل للتحقيق.

ا ـ عند اختبار $H_0: \beta_1 = 5$ ضد $H_a: \beta_1 \neq 5$ بواسطة اختبار خطّي عام، ما هو النموذج المخفض؟ ماعدد درجات الحرية df_R ؟

ب ـ عند اختبار $H_0: \beta_0 = 2$ و $\beta_1 = 5$ ضد ليس كلا من $\beta_0 = 2$ و $\beta_1 = 5$، محققا مستخدمين اختبارا خطيا عاما، ماهو النموذج الناقص؟ ماهي درجات الحرية df_R ؟

(٣ـ٤٤) استنبط (3.75) من (3.71) مستخدما النتيجة في تمرين (٣ـ٤١).

مشاريع

(٣ـ٤٥) بالعودة إلى مجموعة بيانات الـ ***SMSA*** ومشروع (٢ـ٤٢). مستخدما r^2 كمعيار أي متغير مستقل يفسّر أكبر تخفيض في متغيرية عدد الأطباء العاملين؟

(٣-٤٦) بالعودة إلى مجموعة بيانـات الـ **SMSA** والمشـروع (٢-٤٣). أوجـد تقديـرا بفترة لـ β_1 وذلك لكل منطقة على حدة. استخدم 90 بالمائة معامل ثقة في كل حالة. هل تبدو الميول متماثلة في خطوط الانحدار للمناطق المختلفة؟

(٣-٤٧) بالعودة إلى مجموعة بيانات الـ **SENIC** والمشروع (٢-٤٤). ومستخدما r^2 كمعيار، أي المتغيرات المستقلة يفسر أكبر تخفيض في متغيرية معدل طول الإقامة؟

(٣-٤٨) بالعودة إلى مجموعة بيانات الـ **SENIC** ومشـروع (٢-٤٥). أوجـد تقديـرا بفترة لـ β_1 وذلك لكل منطقة على حدة. استخدم 95 بالمائة معامل ثقة لكل حالة. هل تبدو ميول خطوط الانحدار متماثلة في المناطق المختلفة؟

(٣-٤٩) سوف تؤخذ خمس مشاهدات في Y عندمـا $X = 4,\ 8,\ 12,\ 16,\ 20$ وعلـى الـترتيب. ودالـة الانحـدار الحقيقيـة هـي $E\{Y\}=20+4X$ والـ ε_i متغـيرات $N(0,25)$.

ا ـ ولّد خمسة أرقـام طبيعيـة عشـوائية بمتوسـط 0 وتبـاين 25. اعتـبر هـذه الأرقام العشوائية كحدود خطأ للمشاهدات Y الخمس عند $X = 4, 8, 12, 16, 20$ واحسب Y_1، Y_2، Y_3، Y_4 وY_5. أوجد تقديـرات المربعـات الدنيا b_0 و b_1 عند توفيق خط مستقيم للمشاهدات الخمـس. كذلـك احسب $\hat{Y}_h$ عندما $X_h = 10$.

ب ـ أعد الجزء الأول 200 مرة، مولدا أرقاما عشوائية جديدة كل مرة.

جـ ـ ضع توزيعا تكراريا للمائتي تقديرا لـ b_1، واحسب المتوسط والانحراف المعياري للمائتي تقدير لـ b_1. هل تتسق النتائج مع التوقعات النظرية؟.

د ـ لكل من التكرارات الـ 200 احسب 95 بالمائة فترة ثقة لـ $E\{Y_h\}$ عندمـا $X=10$. ما هي نسبة فترات الثقة من بـين الـ 200 فـترة الـتي تتضمـن $E\{Y_h\}$؟ هل هذه النتيجة متسقة مع التوقعات النظرية؟

تشخيصات وتدابير علاجية

عند اختبار نموذج انحدار، مثل نموذج الانحدار الخطّي البسيط (3.1)، لتطبيق ما، فإنه لايمكن، عادة، التأكد مقدما من أن النموذج مناسب لذلك التطبيق. وقد لاتكون سمة أو أكثر من السمات المميزة للنموذج مثل خطّية دالة الانحدار أو طبيعية حدود الخطأ، مناسبة للبيانات التي حصلنا عليها بالذات. وبالتالي فمن المهم فحص صلاحية النموذج للبيانات قبل المضي في مزيد من التحليل الذي يعتمد على ذلك النموذج. ونناقش، في هذا الفصل، بعض الطرق البيانية البسيطة لدراسة صلاحية نموذج، بالإضافة إلى بعض الاختبارات الإحصائية الرسمية للقيام بذلك. ونختم الفصل بمناقشة بعض التقانات التي يمكن أن تجعل نموذج الانحدار الخطّي البسيط (3.1) مناسبا في الوقت الذي لاتتفق البيانات مع شروط النموذج.

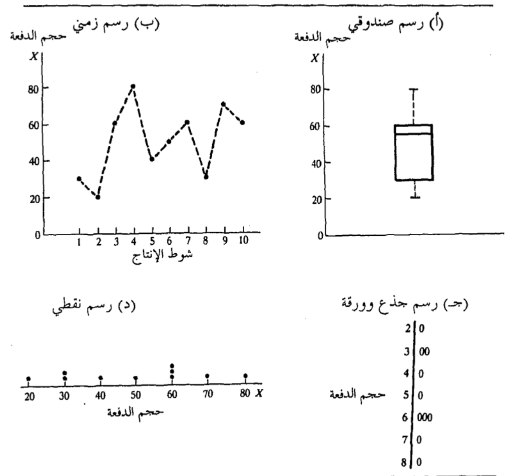
في حين أن المناقشة في هذا الفصل تتطرق لصلاحية نموذج الانحدار الخطّي البسيط (3.1) فإن المبادىء الأساسية المعروضة هنا تنطبق على جميع النماذج الإحصائية المدروسة في هذا الكتاب. وسنستخدم، في فصول لاحقة، مادة إضافية تتعلق بصلاحية نموذج وبتدابير علاجية.

(٤-١) تشخيصات للمتغير المستقل

نبدأ بدراسة بعض التشخيصات البيانية للمتغير المستقل. يحتوي شكل (٤-١)أ رسما صندوقيا بسيطا لحجوم الدفعات في مثال شركة وستوود في الجدول (٢-١). ويبين الرسم الصندوقي في الشكل (٤-١)أ حجمي الدفعة الأصغر والأكبر، الربيعين الأول والثالث وحجم الدفعة الوسيط. ونرى من الشكل (٤-١)أ أنه لاتوجد أحجام دفعات قاصية. وكذلك نرى، وبسبب كون الوسيط قريبا من النهاية العليا للصندوق، أن

الجزء المركزي لتوزيع حجم الدفعات ملتوٍ. ويقدم طولا الخطين المتقطّعين من كل رُبيع إلى النهاية القريبة منه معلومات إضافية عن مخطط حجوم الدفعات. وهنا أيضا تقترح الخطوط المتقطعة بعض الالتواء في الذيل الأيمن.

شكل (٤-١) رسومات تشخيصية للمتغير المستقل ـ مثال شركة وستوود



والتشخيص الثاني المفيد للمتغير هو الرسم الزمني. ويحتوي الشكل (٤-١)ب رسما زمنيا لحجوم الدفعات في مثال شركة وستوود. وقد رسم حجم الدفعة هنا في مقابل شوط الإنتاج (أي، مقابل الزمن) ووصلت النقاط في الرسم لتبيان التتابع الزمني بوضوح أكثر وينبغي الاستفادة من الرسومات الزمنية كلما حصلنا عل بيانات مرتّبة زمنيا. لاتحتوي البيانات في الشكل (٤-١)ب على نمطية خاصة. فلو تبين من الرسم،

مثلا، أن حجوما صغيرة من الدفعات قد استخدمت في البداية وحجوما أكبر فيما بعد، فقد تكون هذه المعلومة مفيدة جدا في الدراسات التشخيصية لصلاحية نموذج الانحدار التوفيقي.

ويحتوي الشكلان (٤-١)جـ و(٤-١)د، رسمين تشخيصيين آخرين يقدّمان معلومات مشابهة لما قدمه الرسم الصندوقي في الشكل (٤-١)أ. إذ يقدم رسم الجذع والورقة في الشكل (٤-١)جـ معلومات مشابهة للمدرج التكراري. وبعرض الأرقام الأخيرة، فإن هذا الرسم يشير أيضا إلى أن كل حجوم الدفعة في مثال شركة وستوود كانت من مضاعفات العشرة.

والرسم النقطي في الشكل (٤-١)د مفيد عندما توجد مشاهدات قليلة فقط في مجموعة البيانات أو عندما يوجد عدد محدود فقط من النتائج في البيانات.

ويشير كل من رسم الجذع والورقة والرسم النقطي، كما أشار الرسم الصندوقي، إلى أن حجوم الدفعات في مثال شركة وستوود ليس متناظر التوزيع تماما. ويظهر الرسمان أيضا أن عدة أشواط قد تمت لحجمي الدفعة 30 و 60.

(٤-٢) الرواسب

عادة، لا تكون الرسوم التشخيصية للمتغير التابع Y مفيدة جدا في تحليل الانحدار لأن قيمة المشاهدات على المتغير التابع دالة في مستوى المتغير المستقل. وبدلا من ذلك تتم تشخيصات المتغير التابع، عادة، بصورة غير مباشرة من خلال فحص الرواسب.

الراسب e_i كما عرفناه في (2.16)، هو الفرق بين القيمة الملحوظة والقيمة التوفيقية:

$$e_i = Y_i - \hat{Y}_i \tag{4.1}$$

وهكذا، يمكن اعتبار الراسب كخطأ ملحوظ، تمييزا له عن الخطأ الحقيقي ε_i في نموذج الانحدار وهو خطأ غير معروف.

$$\varepsilon_i = Y_i - E\{Y_i\} \tag{4.2}$$

وفي نموذج الانحدار (3.1) نفترض أن الـ ε_i متغيرات عشوائية طبيعية مستقلة بمتوسط 0 وتباين ثابت σ^2. وإذا كان النموذج مناسبا للبيانات الملحوظة فينبغي عندئذ

للرواسب التي نحصل عليها e_i أن تعكس الخواص المفروضة لـ ε_i. وهذه هي الفكرة الأساسية التي يستند اليها تحليل الرواسب، الوسيلة المفيدة جدا لفحص صلاحية نموذج.

خواص الرواسب

المتوسط. متوسط الرواسب e_i لنموذج الانحدار الخطّي البسيط (3.1) وعددها هو n بالاستناد إلى (2.17).

$$\overline{e} = \frac{\sum e_i}{n} = 0 \qquad (4.3)$$

حيث $\overline{e}$ تشير إلى متوسط الرواسب. وبما أن $\overline{e}$ تساوي دائما الصفر فإنها لاتعطي معلومات عما إذا كانت القيمة المتوقعة للأخطاء الحقيقية ε_i مساوية للصفر $E(\varepsilon_i) = 0$.

التباين. نعرف تباين الـ n راسبا e_i في نموذج الانحدار (3.1) كما يلي:

$$\frac{\sum (e_i - \overline{e})^2}{n-2} = \frac{\sum e_i^2}{n-2} = \frac{SSE}{n-2} = MSE \qquad (4.4)$$

إذا كان النموذج مناسبا فإن MSE كما لوحظ سابقا، مقدّر غير منحاز لتباين حدود الخطأ σ^2.

عدم الاستقلال. الرواسب e_i ليست متغيرات عشوائية مستقلة لأنها تتضمن القيم التوفيقية $\hat{Y}_i$ والتي ترتكز على تقديري العينة b_0 و b_1. وهكذا يقترن برواسب نموذج الانحدار $n-2$ درجة حرية فقط. وكنتيجة، نعلم من (2.17) أن مجموع e_i الرواسب يجب أن يكون صفرا ومن (2.19) نعلم أن مجموع الجداءات $X_i e_i$ يجب أن يكون صفرا.

وعندما يكون حجم العينة كبيرا بالمقارنة مع عدد المعالم في نموذج الانحدار فإن تأثير عدم استقلال الرواسب e_i غير مهم نسبيا، ويمكن، لمعظم الأغراض، تجاهله.

الرواسب المعيارية.

تُستخدم الرواسب المعيارية أحيانا في تحليل الرواسب. وحيث إن الانحراف المعياري لحدود الخطأ ε_i هو σ ويُقدر بـ $\sqrt{MSE}$، فسوف نعرّف هنا الراسب المعياري كما يلي:

$$\frac{e_i-\bar{e}}{\sqrt{MSE}}=\frac{e_i}{\sqrt{MSE}} \qquad (4.5)$$

وكما سنرى، يمكن أن تكون الرواسب المعيارية مفيدة على وجه الخصوص، في تحديد المشاهدات القاصية.

وإلى جانب الرواسب المعيارية، هناك أيضا مقاييس أخرى مرتكزة على الرواسب ومفيدة في دراسة صلاحية نموذج الانحدار. وسوف نتطرق لها في الفصل الحادي عشر.

انحرافات عن النموذج لدراستها بطريقة الرواسب

سوف ندرس استخدام الرواسب لاختبار ستة أنواع مهمة من الانحرافات عن نموذج الانحدار الخطّي البسيط (3.1) بأخطاء طبيعية:

١- دالة الانحدار ليست خطّية.

٢- حدود الخطأ ليس لها تباين ثابت.

٣- حدود الخطأ ليست مستقلة.

٤- النموذج ملائم لجميع المشاهدات باستثناء مشاهدة واحدة أو قليل من المشاهدات القاصية.

٥- حدود الخطأ ليست طبيعية.

٦- متغير مستقل مهم واحد أو عدد من المتغيرات المستقلة المهمة قد حُذفت من النموذج.

(٤-٣) استخدام الرواسب للتشخيص

نستخدم الآن بعض الرسومات التشخيصية للرواسب لتزودنا بمعلومات عما إذا كان هناك أي من الأنواع الستة للانحراف عن نموذج الانحدار الخطي البسيط المذكورة آنفا.

وسنستخدم هنا لهذا الغرض رسومات الرواسب (أو الرواسب المعيارية) التالية:

١- رسم الرواسب مقابل متغير مستقل.

٢- رسم الرواسب مقابل قيم توفيقية.

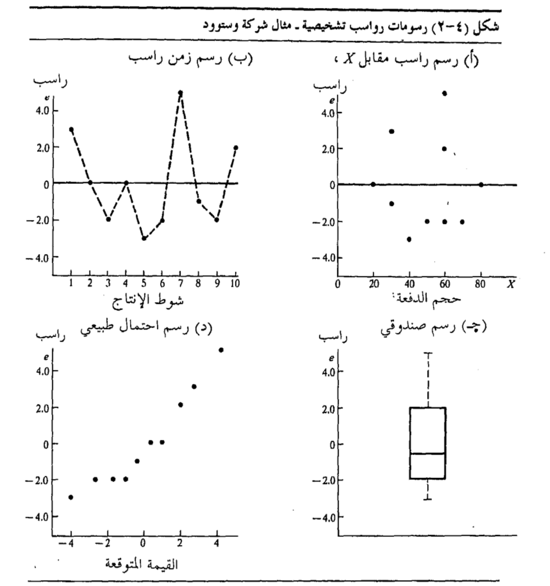
٣- رسم الرواسب مقابل زمن.

٤- رسم الرواسب مقابل متغير مستقل محذوف.

٥- رسم صندوقي للرواسب.

٦- رسم احتمال طبيعي للرواسب.

شكل (٤-٢) رسومات رواسب تشخيصية ـ مثال شركة وستوود



يحوي الشكل (٤-٢)، لمثال شركة وستوود، رسوم الرواسب في الجدول (٢-٣) مقابل المتغير المستقل، مقابل الزمن، رسم صندوقي ورسم احتمال طبيعي. وتدعم كل هذه الرسومات (كما سنرى) صلاحية نموذج الانحدار لبيانات حجوم الدفعات.

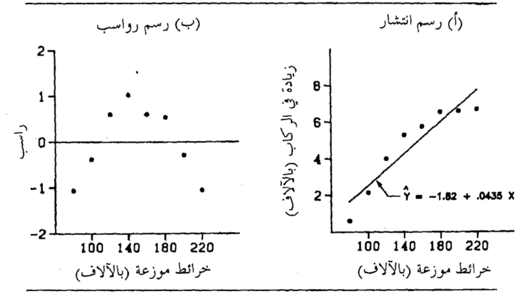
لاخطية دالة الانحدار

يمكن دراسة ما إذا كانت دالة الانحدار الخطّية مناسبة للبيانات قيد التحليل من رسم الرواسب ضد المتغير المستقل أو من رسم الرواسب مقابل القيم التوفيقية وكذلك

من رسم الانتشــار. وليس الأخير، على أي حـال، فعّـالا علـى الـدوام فعاليـة رسم الراسب. ويتضمن الشـكل (٤-٣)أ رسـم انتشـار وخـط انحـدار توفيقـي، مـن إنتـاج الحاسب، للبيانات في دراسة للعلاقـة بـين مقـدار المعلومـات المتعلقـة بـالنقل وركّـاب حافلات النقل في المدن الثماني التي تناولها اختبار مقارنة، حيث X عدد خرائط طرق حافلات نقل الركّاب الموزعة مجانا على سكان المدينة في بداية فترة الاختبار وY مقدار الزيادة، خلال فـترة الاختبـار، في المعـدّل اليومـي لركّـاب الحـافلات في غـير سـاعات الذروة. البيانات الحقيقية والقيم التوفيقية معطـاة في الأعمـدة ١ ، ٢ و٣ مـن الجـدول (٤-١). ويقترح الرسم البياني بقوة عدم ملاءمة دالة انحدار خطّية.

ويقـدم الشـكل (٤-٣)ب للمثـال نفسـه رسـم حاسـب للرواسـب e المبينـة في العمود ٤ من الجدول (٤-١)، مرسومة في مقابل المتغير المسـتقل X وهـو يقـترح أيضـا بقوة عدم ملاءمة دالة الانحدار الخطّية كمـا يتضـح في الشـكل (٤-٣)ب، فالرواسـب تحيد عن 0 بطريقة متناسقة. إذ نلاحظ أنها سالبة من أجل قيم X الصغيرة وموجبة من أجل قيم متوسطة الحجم لـ X وسالبة مرة أخرى من أجل قيم X الكبيرة.

شكل (٤-٣) رسم انتشار ورسم رواسب يوضحان دالة انحدار غير خطية – مثال النقل.



جدول (٤-١) عدد الخرائط الموزّعة وزيادة الركاب ـ مثال النقل

المدينة i	(١) الزيادة في عدد الركاب (بالآلاف) Y_i	(٢) الخرائط الموزعة (بالآلاف) X_i	(٣) القيمة التوفيقية $\hat{Y}_i$	(٤) الراسب $Y_i - \hat{Y}_i = e_i$	(٥) الراسب المعياري $e_i / \sqrt{MSE}$
1	0.60	80	1.66	−1.06	−1.22
2	6.70	220	7.75	−1.05	−1.21
3	5.30	140	4.27	1.03	1.18
4	4.00	120	3.40	0.60	0.69
5	6.55	180	6.01	0.54	0.62
6	2.15	100	2.53	−0.38	−0.44
7	6.60	200	6.88	−0.28	−0.32
8	5.75	160	5.14	0.61	0.70

$$\hat{Y} = -1.82 + 0.0435X$$

$$MSE = 0.756$$

في هذه الحالة كلا الشكلين (٤-٣)أ و(٤-٣)ب وسيلتان فعّالتان في فحص صلاحية خطية دالة الانحدار. وبصورة عامة، على أي حال، يتفوق رسم الراسب على رسم الانتشار ببعض الميزات المهمة. **أولا**، يمكن استخدام الرواسب بسهولة لفحص أوجه أخرى لصلاحية النموذج. **وثانيا**، هناك حالات يمكن فيها لإعادة التدريج في رسم الانتشار أن تجعل المشاهدات Y_i قريبة من القيم التوفيقية $\hat{Y}_i$، كحالة عدم وجود ميل حاد، مثلا. وعندئذ تصبح دراسة صلاحية دالة الانحدار الخطية باستخدام رسم الانتشار أكثر صعوبة. وعلى الوجه الآخر، فإنه يمكن لرسم الراسب، تحت هذه الشروط، أن يبين بجلاء أي نمطية منتظمة في الانحرافات حول خط الانحدار التوفيقي.

ويبين الشكل (٤-٤)أ رسم الراسب مقابل X لحالة نموذجية وذلك عندما يكون النموذج الخطي مناسبا وتنحو الرواسب إلى الوقوع ضمن شريط أفقي متمركز حول 0، دون أن تُظهر اتجاهات منتظمة لأن تكون موجبة وسالبة. وهذه هي الحالة في الشكل (٤-٢)أ لمثال شركة وستوود.

ويبين الشكل (٤-٤)ب حالة نموذجية لانحراف عن نموذج الانحدار الخطّي موضحا الحاجة إلى دالة انحدار منحنية. هنا تنحو الرواسب إلى التغير بصورة منتظمة

بين كونها موجبة وسالبة. وهذه هي الحالة في الشكل (٤-٣)ب لمثال النقل. وبالطبع، سيعود نوع آخر من الانحراف عن الخطّية إلى صورة مختلفة عن النمط النموذجي المبين في الشكل (٤-٤)ب.

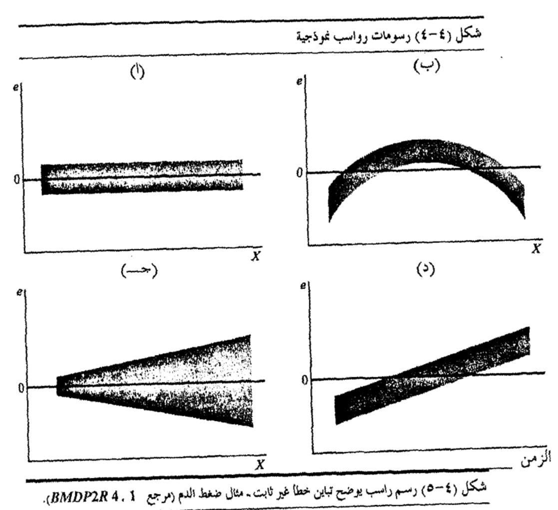
ملاحظة

في نموذج الانحدار الخطّي البسيط، يقدم رسم الرواسب مقابل القيم التوفيقية $\hat{Y}$ معلومات مكافئة لرسم الرواسب مقابل X. والسبب في ذلك هو أن القيم التوفيقية $\hat{Y}_i$ لنموذج الانحدار الخطّي البسيط هي دوال خطّية في القيم X_i للمتغير المستقل؛ وهكذا، يتأثر فقط تدريج القياسات X وليس النسق الأساسي للنقاط المرسومة. ومن أجل الانحدار المنحني والانحدار المتعدد، يكون من المفيد، عادة، رسم الرواسب بصورة منفصلة مقابل القيم التوفيقية ومقابل المتغيرات المستقلة.

عدم ثبات تباين الخطأ

رسومات الرواسب مقابل المتغير المستقل أو مقابل القيم التوفيقية ليست مفيدة فقط في دراسة ما إذا كان النموذج الخطّي مناسبا وإنما كذلك في فحص ما إذا كان تباين حدود الخطأ ثابتا. ويبين الشكل (٤-٥) رسم الراسب في مقابل القيم التوفيقية $\hat{Y}$ في تطبيق يتضمن انحدار ضغط الدم الانبساطي (Y) لأطفال من جنس الإناث مقابل أعمارهن (X) والرسم ناتج عن استخدام حزمة الحاسب *BMDP* (مرجع 4.1). لاحظ أن الخط الأفقي يسمى هنا محور قيم "التنبؤ" وهو مصطلح يستخدم عادة كبديل للقيم التوفيقية. وتشير القيم العددية المبينة في الرسم إلى عدد الرواسب الواقعة عند نقطة ما أو قربها. وأضفنا خطين يتسعان تدريجيا لالقاء الضوء على نزعة أنه كلما كبرت القيم التوفيقية $\hat{Y}$ زاد انتشار الرواسب. وبما أن العلاقة بين ضغط الدم والعمر إيجابية، فإن هذا يقترح أن تباين الخطأ هو أكبر للأطفال الكبار منه للأطفال الصغار.

شكل (٤-٤) رسومات رواسب نموذجية



شكل (٤-٥) رسم راسب يوضح تباين خطأ غير ثابت ـ مثال ضغط الدم (مرجع BMDP2R 4 . 1).

30.0
22.5
15.0
7.50
0.00
-7.50
-15.0

الراسب

61.2 62.1 63.0 63.9 64.8 65.7 66.6 67.5 68.4 69.3

قيمة التنبؤ

الرسم النموذجي في شكل (٤-٤)أ يمثل رسم راسب عندما يكون تباين حد الخطأ ثابتا. ورسم الراسب في شكل (٤-٢)أ لمثال شركة وستوود هو من هذا النوع، مما يقترح هنا أن تباين حدود الخطأ ثابت.

ويوضّح الشكل (٤-٤)جـ صورة نموذجية لرسم راسب يزداد فيه تباين الخطأ مع X. وفي العديد من تطبيقات التجارة، والاقتصاد، علم الاجتماع وعلم الأحياء، تميل الانحرافات عن ثبات تباين الخطأ إلى اتخاذ شكل شبه منحرف كما هو موضح في الشكل (٤-٤)جـ، وكما وجدنا في مثال ضغط الدم في الشكل (٤-٥). ويمكن أن نواجه كذلك تباينات خطأ تتناقص مع زيادة مستويات المتغير المستقل أو تتغير وفق أشكال أخرى التغير.

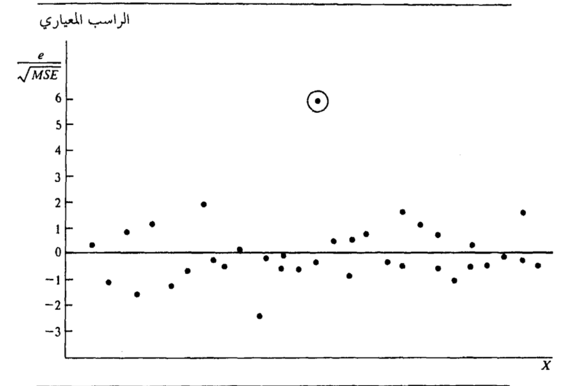
وجود القاصيات

القاصيات هي مشاهدات نائية. ويمكن تحديد الرواسب القاصية من رسومات الرواسب مقابل X أو $\hat{Y}$ (من المفضل أن نستند إلى رواسب معيارية)، وأيضا من رسومات الصناديق ورسومات الجذع والورقة والرسومات النقطية. وفي رسومات الرواسب المعيارية، فإن القاصيات هي نقاط تقع بعيدا عن مواطن انتشار بقية الرواسب وربما كان ذلك بأربعة انحرافات معيارية أو أكثر عن الصفر. ويقدم رسم الراسب في شكل (٤-٦) رواسب معيارية ويتضمن قاصية واحدة، أحيطت بدائرة. لاحظ أن هذا الراسب يمثل مشاهدة تبتعد عن القيمة التوفيقية بستة انحرافات معيارية تقريبا.

ويمكن أن تخلق القاصيات صعوبة كبيرة، وعندما تواجهنا واحدة، فاشتباهنا الأول هو أن هذه المشاهدة نتجت عن غلطة ما أو عن تأثير خارجي وبالتالي ينبغي استبعادها. وأحد الأسباب الرئيسة لاستبعادها هو أنه تحت طريقة المربعات الدنيا، قد ينسحب التوفيقي بصورة غير عادية في اتجاه المشاهدة القاصية ذلك لأننا نريد جعل مربعات الانحرافات أقل مايمكن وقد يسبب هذا توفيقا مضللا، إذا كانت المشاهدة القاصية ناتجة حقا عن غلطة أو عن سبب خارجي. وعلى الوجه الآخر، قد تُبلغنا القاصيات معلومات ذات مغزى، كالحالة التي تقع فيها القاصية بسبب وجود تفاعل

مع متغير مستقل آخر حُـذف مـن النمـوذج، وكثـيرا ماتُقـترح قـاعدة مأمونـة تقضـي باستبعاد قاصية فقط عند وجود دليل مباشر علـى أنهـا تمثـل خطـأ في التسـجيل أو في الحساب، أو سوء استخدام المعدات، أو ظروفا مشابهة.

شكل (٤ ـ ٦) رسم راسب مع قاصية

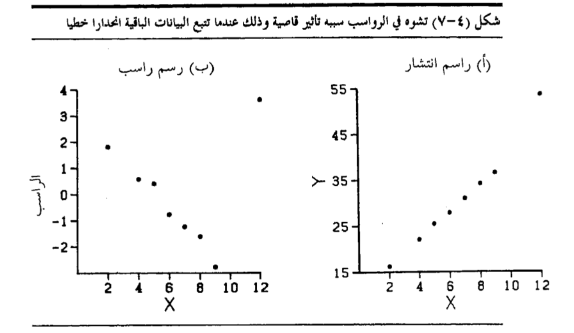


ملاحظة

عند توفيق نموذج انحدار خطّي لمجموعة بيانات بعدد قليل من المشاهدات فيها مشـاهدة قاصية فقد تشوّه المشاهدة القاصية الانحدار التوفيقي إلى حد يقترح فيه رسـم الراسـب نقصـا في توفيـق نمـوذج الانحـدار الخطّي بالإضافـة إلى أنـه يشـير بوضـوح إلى وجـود القاصية. ويوضّح الشكل (٤–٧) هذه الحالة ويقدم رسم الانتشار في الشكل (٤–٧)أ حالة تقع فيها جميع المشاهدات ماعدا القاصية حول علاقة إحصائية علـى شـكل خـط مستقيم. وعند توفيق دالة انحدار خطّية لهذه البيانات، تسبب القاصية إزاحة واضحـة في خط الانحدار التوفيقي تقود إلى نمط منتظم من الانحرافات للمشاهدات الأخـرى عـن الخـط

التوفيقي، كما يوضّح رسم الراسب في الشكل (٤-٧)ب.

شكل (٤-٧) تشوه في الرواسب سببه تأثير قاصية وذلك عندما تتبع البيانات الباقية انحدارا خطيا

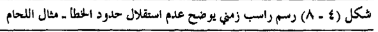


عدم استقلال حدود الخطأ

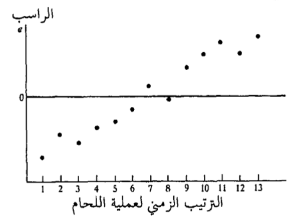
عند الحصول على بيانات وفق تتابع زمني فمن المفيد القيام برسم زمني للرواسب. والغرض من رسم الرواسب مقابل الزمن هو رؤية ما إذا كان يوجد أي ارتباط بين حدود الخطأ مع مضي الزمن. ويحوي الشكل (٤-٨) رسما زمنيا للرواسب في تجربة لدراسة العلاقة بين قطر قطعة ملحومة X وقوة اللحام Y. ويبرز ارتباط واضح بين حدود الخطأ. وتترافق الرواسب السالبة بصورة رئيسة مع المحاولات الأولى والرواسب الموجبة مع المحاولات الأخيرة. وعلى ما يبدو هناك بعض التأثير المرتبط بالزمن، إذ تزداد خبرة عامل اللحام أو تطرأ تغيرات تدريجية في معدات اللحام، لذا تنحو قوة اللحام إلى أن تكون أكبر في اللحامات المتأخرة بسبب هذا التأثير.

وقُدِّم في الشكل (٤-٤)د رسم راسب نموذجي يوضح تأثيرا يعود إلى الزمن وهو يصور تأثيرا يعود إلى الزمن كما في مثال اللحام. ومن المفيد، أحيانا، النظر إلى مسألة

عدم استقلال حدود الخطأ كمسألة حذفنا فيها متغيرا مهما من النموذج (متغير الزمن في حالتنا هنا) وسنناقش هذا النوع من المسائل قريبا.



شكل (٤ - ٨) رسم راسب زمني يوضح عدم استقلال حدود الخطأ - مثال اللحام



عندما تكون حدود الخطأ مستقلة، نتوقع تذبذب الرواسب بشكل أو بآخر تذبذبا عشوائيا حول خط الأساس 0: مثل الانتشار المبين في الشكل (٤-٢)ب لمثال شركة وستوود. ويتخذ نقص العشوائية شكل زيادة حادة في تذبذب النقاط حول الخط الصفري، أو شكل تذبذب ضعيف جدا حوله. وفي التطبيقات لا نهتم كثيرا بالحالة الأولى لأنها لا تظهر كثيرا. وعلى النقيض، فإن التذبذب الضعيف متواتر الحدوث، كما في مثال اللحام في الشكل (٤-٨).

ملاحظة

عند رسم الرواسب مقابل X، كما في شكل (٤-٣)ب قد لا يبدو الانتشار عشوائيا. وفي هذا الرسم، قد لا تكون المسألة الأساسية، على أي حال، هي ضعف استقلالية حدود الخطأ وإنما ضعف توفيق دالة الانحدار. وهذه هي، في الحقيقة، الحالة التي يصورها رسم الانتشار في الشكل (٤-٣)أ.

لاطبيعية حدود الخطأ

كما لاحظنا سابقا، لا تسبب الانحرافات الطفيفة عن الطبيعية مشاكل جدية.

ومن جهة أخرى، ينبغي أن تكون الانحرافات الكبيرة موضع الاهتمام، ويمكن دراسة طبيعية حدود الخطأ دراسة غير رسمية، وذلك بفحص الرواسب مستخدمين تشكيلة من الطرق البيانية.

رسومات توزيع. رسم الصندوق مفيد للحصول على معلومات ملخصة عن تناظر الرواسب وعن قاصيات محتملة. ويحوي الشكل (٤-٢)جـ رسم الصندوق للرواسب في مثال شركة وستوود. ولا يقترح هذا الرسم وجود انحرافات كبيرة عن الطبيعية. ويمكن إقامة مدرج تكراري أو رسم نقطي أو رسم جذع وورقة للرواسب وذلك لرؤية ما إذ اكانت هذه الرسومات تشير إلى انحرافات كبيرة عن الطبيعية. وعلى أي حال، ينبغي أن يكون عدد المشاهدات في دراسة الانحدار كبيرا نسبيا كي يمكن لأي من هذه الرسومات أن يقدم معلومات موثوقة عن شكل توزيع حدود الخطأ.

مقارنة التكرارات. والإمكانية الأخرى هي مقارنة التكرارات الفعلية للرواسب مع التكرارات المتوقعة تحت الطبيعة، فمثلا، إذا كان عدد المشاهدات في دراسة الانحدار كبيرا إلى حد ما، فيمكن تحديد ما إذا كان حوالي 68 بالمائة من الرواسب المعيارية $e_i/\sqrt{MSE}$ مثلا، واقعا بين 1 - و1 أو ما يقارب من 90 بالمائة من الرواسب المعيارية واقعا بين 1.645 - و 1.645. وإذا كان حجم العينة صغيرا، فيمكن استخدام قيمة t المقابلة للمقارنة.

ولتوضيح هذه الطريقة، نعتبر مرة أخرى مثال النقل في الجدول (٤-١). إذ يحوي العمود رقم (٥) الرواسب المعيارية التي حصلنا عليها بتقسيم كل راسب في العمود (٤) على $\sqrt{MSE}=\sqrt{0.756}$. وبما أن n صغيرة فسوف نستخدم للمقارنة توزيع t بـ $n-2=8-2=6$ درجات حرية. والمئين الـ 95 لهذا التوزيع هو 1.943 من جدول (أ-٢) ، وهكذا نتوقع تحت الطبيعية أن يقع حوالي 90 بالمائة من الرواسب المعيارية بين 1.943 - و1.943. وهنا، تقع كل الرواسب المعيارية الثمانية ضمن هذين الحدين. وبالمثل نتوقع تحت الطبيعية، أن يقع حوالي 60 بالمائة من الرواسب المعيارية بين

-0.906 و0.906. والنسبة الفعلية هنا هي 62.5 بالمائة. وهكذا تتسق التكرارات الفعلية هنا مع تلك المتوقعة تحت الطبيعية اتساقا معقولا.

رسم الاحتمال الطبيعي. والإمكانية الأخرى أيضا، هي إعداد رسم طبيعي للرواسب، فنرسم هنا كل راسب مقابل قيمته المتوقعة عندما يكون التوزيع طبيعيا. والرسم الذي يكون خطيا تقريبا يقترح اتفاقا مع الطبيعية، بينما يقترح الرسم الذي ينحرف بصورة ملموسة عن الخطية أن توزيع الخطأ ليس طبيعيا.

ويحوي العمود ١ من الجدول (٤-٢) الرواسب لمثال شركة وستوود وبترتيب تصاعدي (من جدول ٢-٣). ولإيجاد القيم المتوقعة للرواسب المرتبة تحت الطبيعية، نستخدم الحقائق التالية: (١) القيم المتوقعة لحدود الخطأ لنموذج الانحدار (3.1) هي صفر، و(٢) الانحراف المعياري لحدود الخطأ يقدر بـ $\sqrt{MSE}$. وتبين نظرية الإحصاء أنه في حالة متغير عشوائي طبيعي متوسطة 0 وانحرافه المعياري مقدر بـ $\sqrt{MSE}$، يكون التقريب التالي:

$$MSE\left[z\left(\frac{i-0.375}{n+0.25}\right)\right] \tag{4.6}$$

تقريبا جيدا للقيمة المتوقعة للمشاهدة الأصغر الـ i من عينة عشوائية حجمها n، حيث يرمز $z(A)$ كالعادة، للمئين $100(A)$ للتوزيع الطبيعي المعياري.

باستخدام هذا التقريب، دعنا نحسب القيم المتوقعة تحت الطبيعية للرواسب المرتبة لمثال شركة وستوود. فقد وجدنا سابقا (جدول ٣-٣) أن $MSE = 7.5$.

ومن أجل أصغر راسب، لدينا $i = 1$. وبالتالي :

$$(i -0.375) / (n + 0.25) = (1 - 0.375) / (10 +0.25) = 0.061$$

والقيمة المتوقعة لأصغر راسب تحت الطبيعية هي:

$$\sqrt{7.5}\,[z(.061)]=\sqrt{7.5}\,(-1.55)=-4.24$$

وبالمثل، نحصل على القيمة المتوقعة، تحت الطبيعية، لثاني أصغر راسب، بأن نحسب من أجل $i = 2$ ، $(i -.375)/(n +.25)=(2 - .375)/(10 + .25) =.159$ وبالتالي:

جدول (٤-٢) الرواسب والقيم المتوقعة تحت الطبيعية لمثال شركة وستوود.

ترتيب تصاعدي i	(١) الرواسب مرتبة e_i	(٢) القيمة المتوقعة تحت الطبيعية
1	−3.0	−4.24
2	−2.0	−2.74
3	−2.0	−1.79
4	−2.0	−1.02
5	−1.0	−0.33
6	0.0	0.33
7	0.0	1.02
8	2.0	1.79
9	3.0	2.74
10	5.0	4.24

$$\sqrt{7.5}\left[z(.159)\right]=\sqrt{7.5}\,(-1.00)=-2.74$$

وكنتيجة لتناظر التوزيع الاحتمالي الطبيعي، فإن القيمتين المتوقعتين لأكبر وثاني أكبر راسب هما 4.24 و 2.74 ، على الترتيب.

ويحوي العمود ٢ من الجدول (٤-٢) كل القيم العشر المتوقعة تحت فرضية الطبيعية. ويقدم الشكل (٤-٢)د رسم الرواسب مقابل قيمها المتوقعة تحت الطبيعية. ونلاحظ أن النقاط في الشكل (٤-٢)د تقترب اقترابا معقولا من خط مستقيم، مما يشير إلى أن توزيع حدود الخطأ لا ينحرف انحرافا جوهريا عن التوزيع الطبيعي. وتعود الدرجات في الرسم في الشكل (٤-٢)د إلى الطبيعة التقريبية للبيانات في مثال شركة وستوود.

ويقدم العديد من حزم الحاسب الآلي رسوم احتمال طبيعية وفقا لاختيار مستخدميها. وتستخدم بعض هذه الرسومات الرواسب المعيارية ولكن هذا لا يؤثر في الطبيعة الأساسية للرسم.

وبالإضافة إلى تقويم التقريب الخطي للنقاط المرسومة في رسم احتمال طبيعي بالعين المجردة، يمكن كذلك حساب معامل الارتباط (3.73) بين الرواسب وقيمها المتوقعة تحت فرض الطبيعية، والقيم العالية لمعامل الارتباط مؤشر للطبيعية. ويحوي

الجدول (٤-٣) قيما حرجة (معينـات) لتوزيـع معـامل الارتبـاط بـين الرواسـب المرتبـة وقيمها المتوقعة تحت الطبيعية، أي عندما تكون حدود الخطأ متوزعة طبيعيا، وذلك من أجل حجوم مختلفة للعينة. وعند قيمة لـ α، إذا كانت القيمة الملحوظة لمعامل الارتبـاط لا تقل عـن القيمـة المبينـة في الجـدول فيمكـن أن نسـتنتج أن فـرض التوزيـع الطبيعـي كتوزيع لحدود الخطأ هو فرض معقول. وفي مثال شركة وستوود في الجدول (٤-٢)، وجدنا معـامل الارتبـاط 0.955 وبتقييـد المخـاطرة α عنـد 0.05 نشـاهد مـن الجـدول (٤-٣) أن القيمة المقابلة لـ $n = 10$ هي 0.918. وحيـث إن المعـامل الملحـوظ يتخطـى هذا المستوى، فهناك ما يدعم استنتاجنا السابق بأن توزيع حدود الخطأ لا تحيـد كثـيرا عن التوزيع الطبيعي.

ويتضمن الشكل (٤-٩) رسم احتمال طبيعي في دراسة انحدار تتبع حدود الخطـأ فيها توزيعا ملتويا بحدة : وهذا الرسم ناتج عـن اسـتخدام حزمـة مينيتـاب الإحصائيـة (مرجع 4.2). لاحظ الانحراف الكبير عـن الوضـع الخطّـي للنقـاط في شـكل (٤-٩). ومعامل الارتباط بين الـ 14 راسبا المرتبة وقيمها المتوقعة تحت الطبيعية هو 0.88 فقـط، وهـو يشـير كذلـك إلى انحـراف عـن التوزيـع الطبيعـي لأن القيمـة الحرجـة مـن أجـل $\alpha = 0.05$ و $n = 14$، طبقا للجدول (٤-٣)، أكبر من 0.92.

جدول (٤-٣) قيم حرجة لمعامل الارتباط بين رواسب مرتبة وقيمها المتوقعة تحت الطبيعية وذلـك عندما يكون توزيع حدود الخطأ طبيعيا.

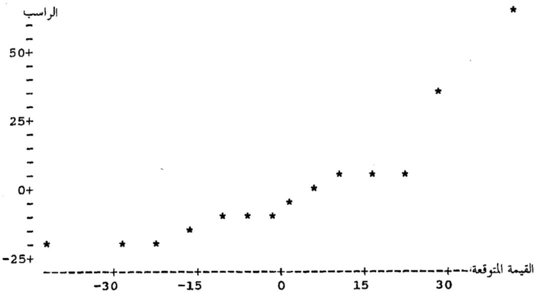
.01	.05	.10	n
مستوى المعنوية α			
.826	.880	.903	5
.879	.918	.934	10
.910	.939	.951	15
.926	.951	.960	20
.939	.959	.966	25
.947	.964	.971	30
.959	.972	.977	40
.966	.977	.981	50
.976	.984	.987	75
.982	.989	.989	100

المصدر: Reprinted, with permission, from S.W. Looney and T.R. Gulledge, Jr., "Use of the Correlation Coefficient with Normal Probability Plots", *The American Statistician* 39 (1985), pp. 75-79.

ملاحظة

التحليل المتعلق بانحرافات النموذج عن الطبيعية، هو من عدّة وجوه، أكثر صعوبة من ذلك المتعلق بأنواع أخرى من الانحرافات. ففي المقام الأول يمكن أن يكون التغير العشوائي مصدر أذى، خصوصا عند دراسة طبيعة توزيع احتمالي، ما لم يكن حجم العينة كبيرا تماما. والأكثر سوءا أن أنواع الانحرافات الأخرى تستطيع بالفعل التأثير في

شكل (٤-٩) مثال رسم احتمال طبيعي عندما يكون توزيع حدود الخطأ ملتويا بحدة
(مرجع 4.2 ، مينيتاب)



توزيع الرواسب. فمثلا، قد تبدو الرواسب غير متوزعة طبيعيا لأن دالة الانحدار المستخدمة غير مناسبة، أو لأن تباين الخطأ غير ثابت، وبالتالي فإن تقصي هذه الأنواع الأخرى من الانحرافات أولا، قبل الاهتمام بطبيعية حدود الخطأ، هو، في العادة، استراتيج جيد.

حذف متغيرات مستقلة مهمة

ينبغي أيضا رسم الرواسب مقابل المتغيرات المحذوفة من النموذج والتي يمكن أن يكون لها تأثيرات مهمة على الاستجابة، وذلك عند توافر البيانات. ومتغير الزمن المذكور سابقا في تطبيق اللحام مثال على ذلك، وغرض هذا التحليل الإضافي هو

تحديد ما إذا كانت هناك أية متغيرات مستقلة رئيسة أخرى يمكن أن تمنح النموذج قوة مهمة في مجال الوصف والتنبؤ.

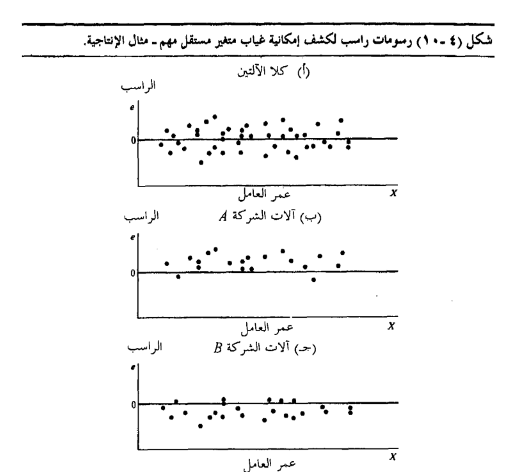
وكمثال آخر، في دراسة للتنبؤ بإنتاجية عامل وفقا لمعدل القطع التي ينتجها في عملية تجميعية، دُرست العلاقة بين إنتاجية العامل Y وعمره X وذلك في عينة من العمال. ورسم الرواسب مقابل X مبين في الشكل (٤ـ١٠)أ ، وهو يشير إلى أنه لا أساس للشك في صلاحية خطّية دالة الانحدار أو ثبات تباين الخطأ.

والآلات المستخدمة في عملية التجميع هي من إنتاج شركتين (A وB). وقد رسمت الرواسب مقابل X حسب نوع الآلة المستخدمة وهي معروضة في الشكلين (٤ـ١٠)ب و(٤ـ١٠)جـ. لاحظ أن الرواسب للآلات المصنوعة في الشركة A تنحو إلى أن تكون موجبة، في حين تنحو تلك الخاصة بآلات مصنوعة في الشركة B إلى أن تكون سالبة. وهكذا، يبدو أن لنوع الآلة تأثيرا مؤكدا في الإنتاجية، وقد تتمخض تنبؤات الإنتاجية عن كونها أفضل بكثير عند إضافة هذا المتغير المستقل إلى النموذج. وفي حين يعالج هذا المثال متغيرا تصنيفيا (نوع الآلة) فإن تحليل الراسب لمتغير كمي إضافي هو تحليل مشابه تماما. وتُرسم الرواسب ببساطة مقابل المتغيرات العشوائية الإضافية ثم ننظر فيما إذا كانت الرواسب تنحو إلى التغير بصورة نمطية مع مستوى المتغير المستقل الإضافي أم لا.

ملاحظة

لا نقول أن النموذج الأصلي "**خاطئ**" عندما يكون من الممكن تحسينه بصورة ملموسة بإضافة متغير مستقل أو أكثر. وفي حالات من عالم الواقع لا يمكن أن يظهر في نموذج الانحدار إلا قليل من العوامل المؤثرة في متغير تابع Y. ولذلك فإن الغرض الرئيس لتحليل الراسب في مجال تحديد متغيرات مستقلة أخرى مهمة هو اختبار كفاية النموذج والنظر في ما إذا كان يمكن تحسينه بصورة ملموسة بإضافة متغير مستقل أو عدد قليل من المتغيرات المستقلة.

شكل (٤ ـ ١٠) رسومات راسب لكشف إمكانية غياب متغير مستقل مهم ـ مثال الإنتاجية.



بعض التعليقات الختامية

١ـ ناقشنا انحرافات النموذج كلا على حدة. وفي الواقع العملي، قد يقع عديد من الانحرافات معا. فمثلا، قد تكون دالة الانحدار الخطية توفيقا فاشلا للبيانات وقد لا يكون تباين حدود الخطأ ثابتا. وفي حالات كهذه يمكن أن تبقى الأنماط النموذجية في الشكل (٤ـ٤) مفيدة، إلا أنها قد تحتاج إلى خلطها في أنماط مركّبة.

٢ـ في حين أن التحليل البياني للرواسب هو مجرد طريقة تحليل غير رسمية، إلا أنها في كثير من الحالات تكفي للتحقق من صلاحية النموذج.

٣- لا ينحصر تطبيــق الطريقـة الأساسـية لتحليـل الرواسـب في نموذج الانحـدار الخطّي البسيط ولكنه ينطبق أيضا على انحـدار أكـثر تعقيـدا وعلـى أنـواع أخـرى مـن النماذج الإحصائية.

٤- يمكن القيام بمعظـم العمـل الروتيـني في تحليـل الرواسـب باسـتخدام الحاسـب الآلي. وتقدم معظم برامج الانحدار، تقريبا، القيم التوفيقية والرواسب المقابلـة، وتتوافـر عموما روتينات يمكن بواسطتها الحصول على أنواع مختلفة من رسومات الراسب.

(٤ - ٤) نظرة إجمالية لاختبارات تتعلق بالرواسب

التحليل البياني للرواسب هو في الأصل تحليل ذاتـي. ومـع ذلـك فـإن التحليـل الذاتـي لأنواع من رسومات الرواسب ذات الصلة ببعضها البعض يكشـف عـن صعوبـات في النموذج بصورة أكثر وضوحا من اختبارات رسمية معينة. ولكن هناك مناسبات، علـى أي حال، نرغب فيها وضع تساؤلات محددة موضع الاختبار. وسنعرض الآن باختصار بعض الاختبارات ذات العلاقة، ونتابع بالتفصيل اختبارا من نوع جديد.

وتتطلب معظم الاختبارات الإحصائية مشاهدات مستقلة. إلا أن الرواسب كمـا رأينا، غير مستقلة. ولحسـن الحـظ، فـإن عـدم الاسـتقلالية تصبـح ضعيفـة في العينـات الكبيرة مما يسمح عادة بتجاهلها.

اختبارات العشوائية

كثيرا ما يُستخدم اختبار الأشـواط لاختبـار نقـص العشـوائية في رواسـب مرتبـة زمنيا. واختبار آخر مصمم خصيصا لنقص العشوائية في رواسب المربعات الدنيا هو اختبـار دربن ـ واطسون (*Durbin - Watson*) ويُناقش هذا الاختبار في الفصل الثالث عشر.

اختبارات ثبات التباين

عندما يُعطي رسم راسب الانطبـاع بـأن التبـاين قـد يكـون مـتزايدا أو متناقصـا بصورة نمطية، بالنسبة لـ X أو لـ $E\{Y\}$، فهناك اختبـار بسـيط ينشـأ عـن توفيـق دالـتي انحدار لكل من نصفي المشاهدات مرتبة وفق مستوى X ثم حساب متوسطي مربعـات الخطأ لكل منها، ثم اختبار تساوي تبايني الخطأ باستخدام الاختبار F. واختبار بسـيط

آخر هو بواسطة ارتباط الرتب بين القيمة المطلقة للراسب وقيمة المتغير المستقل.

اختبارات للقاصيات

ينطوي اختبار بسيط يتعلق بمشاهدة قاصية على توفيق خط انحدار جديد للمشاهدات الـ $n-1$ الأخرى. والآن يمكن اعتبار المشاهدة المشبوهة والتي لم تُستخدم في توفيق الخط الجديد كمشاهدة جديدة، ويمكن في حالة n من المشاهدات نحصل بالمصادفة على انحراف عن خط الانحدار التوفيقي في حجم انحراف المشاهدة القاصية. وإذا كان الاحتمال صغيرا بما فيه الكفاية فيمكن رفض القاصية واعتبارها لم تأت من ذات المجتمع الذي جاءت منه المشاهدات الأخرى الـ $n-1$، وفيما عدا ذلك نحتفظ بالقاصية.

وطُورت اختبارات عديدة أخرى للمساعدة في تقويم المشاهدات القاصية. وقد نوقشت هذه الاختبارات في مراجع متخصصة مثل المرجع [4.3] وفي المجلات الإحصائية.

اختبارات للطبيعية

يمكن استخدام اختبارات جودة التوفيق لاختبار طبيعية حدود الخطأ. فمثلا يمكن استخدام اختبار مربع كاي أو اختبار كولموجوروف ـ سمير نوف *Kolmogorov Simirnov* وتعديله اختبار ليليفورز *Lilliefors* لاختبار طبيعية حدود الخطأ بواسطة تحليل الرواسب.

ملاحظة

اختبار الأشواط، واختبارات ارتباط الرتب وجودة التوفيق هي طرق إحصائية شائعة الاستخدام ومدروسة في العديد من كتب الإحصاء المدرسية الأساسية.

(٤ ـ ٥) اختبار F لنقص التوفيق

ونتابع الآن اختبارا رسميا لتحديد ما إذاكانت دالة انحدار محددة تتوافق بصورة طيبة مع البيانات. ونوضّح هذا الاختبار الذي يهدف إلى التحقق مما إذا كانت دالة انحدار خطّية توفيقا جيدا للبيانات.

الفرضيات

يفترض اختبار نقص التوفيق أن المشاهدات Y المقابلة لـ X معطاة هي: (١)

مستقلة، (٢) متوزعة طبيعيا، وأن (٣) لتوزيعات الـ Y التباين σ^2 نفسه.

وكذلك يتطلب اختبار نقص التوفيق تعدد المشاهدات عند مستوى واحد أو أكثر لـ X. وفي البيانات غير التجريبية، يمكن أن تحدث هذه بالمصادفة، كما في دراسة الإنتاجية التي تربط بين إنتاجية العمال وأعمارهم، ويتفق أن تتضمن الدراسة العديد من العمال في العمر نفسه. أما في تجربة فيمكن التأكد عن طريق التصميم من وجود مشاهدات معادة. فمثلا في تجربة على تأثير حجم عمولة البائع على المبيعات يمكن تقديم عمولة من حجم معين لثلاثة بائعين وذلك، لكل من ستة حجوم للعمولة ثم تُلحظ مبيعاتهم.

تسمى إعادة المحاولات عند المستوى نفسه للمتغير المستقل، من النوع الذي وصفناه، تكرارات، وتُسمى المشاهدات الناتجة متكررات.

مثال

في تجربة تتضمن 12 من المكاتب الفرعية لمصرف تجاري، متشابهة ولكنها متفرقة في الضواحي، عُرضت هدايا على أصحاب الحسابات الجارية في المكاتب لكي يفتحوا حسابات توفير. وينبغي للإيداع الأول في حساب التوفير الجديد أن يتجاوز حدا أدنى محددا للحصول على الهدية. وتتناسب قيمة الهدية مباشرة مع الحد الأدنى للإيداع. واستخدمت مستويات متقدمة للإيداع الأدنى وقيمة الهدية المرتبطة به في التجربة للتحقق من العلاقة بين الإيداع الأدنى المحدد وقيمة الهدية من جهة وعدد الحسابات المفتوحة في المكتب من جهة أخرى، وبالإجمال، استخدمت ستة مستويات من الإيداعات الدنيا وقيم الهدايا المتناسبة معها، وخصص مكتبان فرعيان عشوائيا لكل مستوى، وقد شب حريق في أحد المكاتب خلال فترة الاختبار وأسقط من الدراسة، ويحوي جدول (٤-٤) النتائج حيث X قيمة الإيداع الأدنى و Y عدد حسابات التوفير الجديدة التي افتتحت والمؤهلة للهدية خلال فترة الاختبار.

وفقت دالة انحدار **خطية** بالطريقة المعتادة : وهي (الحسابات غير موضحة).

$$\hat{Y} = 50.72251 + 0.48670X$$

جدول (٤ - ٤). بيانات لمثال البنك

الفرع i	حجم الإيداع الأدنى (بالدولار) X_i	عدد الحسابات الجديدة Y_i	الفرع i	حجم الإيداع الأدنى (بالدولار) X_i	عدد الحسابات الجديدة Y_i
1	125	160	7	75	42
2	100	112	8	175	124
3	200	124	9	125	150
4	75	28	10	200	104
5	150	152	11	100	136
6	175	156			

وتمّ كذلك الحصول على جدول تحليل التباين وهو معروض في الجدول (٤ـ٥)أ. ويبين الشكل (٤ـ١١) رسم انتشار مع خط الانحدار التوفيقي. كانت المؤشرات قوية إلى أن دالة الانحدارالخطّية غير مناسبة. ولاختبار هذا رسميا، سوف نستخدم أسلوب الاختبار الخطي العام المُوصوف في الفقرة (٣ـ٩).

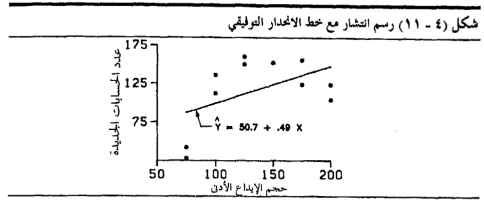
جدول (٤ - ٥) جدول تحاين لمثال المصرف

مصدر التغير	SS	df	MS
الانحدار	$SSR = 5,141.3$	1	$MSR = 5,141.3$
الخطأ	$SSE = 14,741.6$	9	$MSE = 1,638.0$
المجموع	$SSTO = 9,882.9$	10	

رموز

نحتاج إلى تعديل رموزنا لتمييز وجود تكرارات عند بعض مستويات X. ويقدم الجدول (٤ـ٦) البيانات نفسها كما في الجدول (٤ـ٤) ولكن بترتيب يميز المتكررات وسوف نشير إلى مستويات X المختلفة في الدراسة سواء وجدت مشاهدات متكررة أم

شكل (٤ - ١١) رسم انتشار مع خط الانحدار التوفيقي



لا، بـ X_1، X_2،...،X_c. وفي مثال المصرف $c = 6$، حيث يوجد في الدراسة ستة مستويات لحجم الإيداع الأدنى، وفي خمسة منها توجد مشاهدتان، ولواحدة توجد مشاهدة واحدة. وسوف نجعل $X_1 = 75$ (أصغر مستوى إيداع أدنى)، $X_2 = 100$، ....، $X_6 = 200$. إضافة إلى أننا سنرمز لعدد المتكررات عند المستوى j من مستويات X بـ n_j في مثالنا $n_1 = n_2 = n_3 = n_5 = n_6 = 2$ و $n_4 = 1$. وهكذا يكون العدد الكلي للمشاهدات n معطى بالعلاقة:

$$n = \sum_{j=1}^{c} n_j \tag{4.7}$$

وسنرمز للقيمة الملحوظة للمتغير التابع من أجل التكرار i عند المستوى j لـ X بـ Y_{ij} حيث $i = 1,\dots, n_j$ و $j = 1,\dots, c$. وفي مثال المصرف (جدول ٤-٦) لدينا: $Y_{11} = 28$, $Y_{21} = 42$, $Y_{12} = 112$ وهكذا، وأخيرا سوف نشير إلى متوسط مشاهدات Y عند المستوى $X = X_j$ بـ $\overline{Y}_j$. وهكذا $\overline{Y}_1 = \frac{28+42}{2} = 35$ و $\overline{Y}_4 = 152/1 = 152$.

جدول (٤ - ٦) بيانات مثال المصرف، مصنفة وفقا لرقم التكرار والإيداع الأدنى.

التكرار	حجم الإيداع الأدنى (بالدولار)					
	$j = 1$ $X_1 = 75$	$j = 2$ $X_2 = 100$	$j = 3$ $X_3 = 125$	$j = 4$ $X_4 = 150$	$j = 5$ $X_5 = 175$	$j = 6$ $X_6 = 200$
$i = 1$	28	112	160	152	156	124
$i = 2$	42	136	150	.	124	104
المتوسط $\overline{Y}_j$	35	124	155	152	140	114

نموذج تام

يبدأ الاختبار الخطّي العام بتحديد النموذج التام. ويقوم النموذج التام المستخدم لاختبار نقص التوفيق على فرضيات نموذج الانحدار الخطّي البسيط (3.1) نفسها، باستثناء افتراض علاقة انحدار خطّية، وهي موضع الاختبار. وهذا النموذج التام هو:

$$Y_{ij} = \mu_j + \varepsilon_{ij} \qquad (4.8)$$ نموذج تام

حيث μ_j معالم $j = 1,..., c$

ε_{ij} مستقلة و $N(0,\sigma^2)$

وبما أن توقع حدود الخطأ يساوي الصفر، فنستنتج:

$$E\{Y_{ij}\} = \mu_j \qquad (4.9)$$

وهكذا فإن المعلمة $\mu_j (j=1,...,c)$ هي متوسط الاستجابة عند $X = X_j$.

ويشابه النموذج التام (4.8) نموذج الانحدار (3.1) بالتصريح بأن كل استجابة لها تتكون من مركّبتين متوسط الاستجابة عند $X = X_j$ وحد خطأ عشوائي. والاختلاف بين النموذجين هو أنه في النموذج التام (4.8) لا توجد قيود على المتوسطات μ_j، بينما ترتبط متوسطات الاستجابة في نموذج الانحدار (3.1)، ارتباطا خطيا بـ X (أي $E\{Y\} = \beta_0 + \beta_1 X$).

ولتوفيق النموذج التام للبيانات، نحتاج إلى مقدرات المربعات الدنيا للمعالم μ_j . يمكن إثبات أن مقدرات المربعات الدنيا لـ μ_j هي ببساطة متوسطات العينة $\overline{Y}_j$.

$$\hat{\mu}_j = \overline{Y}_j \qquad (4.10)$$

وهكذا فإن القيمة المتوقعة المقدَّرة للمشاهدة Y_{ij} هي $\overline{Y}_j$ ومجموع مربعات الخطأ للنموذج التام هو تبعا لذلك:

$$SSE(F) = \sum_j \sum_i (Y_{ij} - \overline{Y}_j)^2 = SSPE \qquad (4.11)$$

وفي سياق اختبار نقص التوفيق، يسمى مجموع مربعات الخطأ للنموذج التام مجموع مربعات الخطأ البحت ويشار له بـ $SSPE$.

ونلاحظ أن $SSPE$ يتكون من مجموع مربعات الانحرافات عند كل مستوى X.
وعند المستوى $X = X_j$ يكون هذا المجموع للانحرافات المربعة:

$$\sum_i (Y_{ij} - \overline{Y}_j)^2 \qquad (4.12)$$

ومن ثَمَّ تُضاف مجموعات المربعات هذه فوق جميع مستويات X $(j = 1,...,c)$ وفي مثال المصرف، لدينا:

$$\begin{aligned} SSPE = &(28\text{-}35)^2 + (42 - 35)^2 + (112\text{-}124)^2 + (136\text{-}124)^2 \\ &+ (160\text{-}155)^2 + (150\text{-}155)^2 + (152\text{-}152)^2 \\ &+ (156\text{-}140)^2 + (124\text{-}140)^2 + (124\text{-}114)^2 \\ &+ (104\text{-}114)^2 = 1{,}148 \end{aligned}$$

ونلاحظ أن أي مستوى X لا يوجد عنده تكرارات لا يساهم في $SSPE$ لأن $\overline{Y}_j = Y_{1j}$ عندئذٍ. وهكذا فإن $0 = (152 - 152)^2$ من أجل $j = 4$ في مثال المصرف.

ويمكن إيجاد درجات الحرية الموافقة لـ $SSPE$ بمعرفة أن مجموع الانحرافات المربعة (4.12) عند أي مستوى معطى X، مثله مثل مجموع مربعات كلي عادي مبني على n مشاهدة، إذ يوافقه عندئذٍ $n - 1$ درجة حرية. وهنا، يوجد n_j مشاهدة عندما $X = X_j$ وبالتالي فإن درجات الحرية الموافقة هي $n_j - 1$. وكما أن $SSPE$ هو مجموع مجاميع المربعات (4.12)، فإن عدد درجات الحرية الموافقة لـ $SSPE$ هو مجموع درجات الحرية لكل من المركّبات.

$$df_F = \sum_j (n_j - 1) = \sum_j n_j - c = n - c \qquad (4.13)$$

وفي مثال المصرف، لدينا $df_F = 11 - 6 = 5$. لاحظ أن أي مستوى X بدون تكرارات لا يسهم في df_F لأن $n_j - 1 = 1 - 1 = 0$ عندئذٍ، وعلى غرار ذلك فإن مثل هذا المستوى X ليس له أي مساهمة في $SSPE$.

نموذج مخفض

وتتطلب طريقة الاختبار الخطّي العام بعد ذلك اعتبارالنموذج المخفض تحت H_0.
ولاختبار صلاحية علاقة الانحدار الخطّي يكون البديلان:

$$\begin{aligned} &H_0: E\{Y\} = \beta_0 + \beta_1 X \\ &H_a: E\{Y\} \neq \beta_0 + \beta_1 X \end{aligned} \qquad (4.14)$$

وهكذا، تفترض H_0 أن μ_j في النموذج التام (4.8) مرتبطة خطّيا مع X:

$$\mu_j = \beta_0 + \beta_1 X_j$$

ولذلك يكون النموذج المخفض:

$$Y_{ij} = \beta_0 + \beta_1 X_j + \varepsilon_{ij} \qquad \text{نموذج مخفض} \tag{4.15}$$

ونلاحظ أن النموذج المخفض هو نموذج الانحـدار الخطّي البسـيط المعتـاد (3.1) بأدلة معدّلة لتمييز وجود تكرارات. ونعلم أن القيم المتوقعة المقدَّرة للمشـاهدة Y_{ij} تبعـا لنموذج الانحدار (3.1) هي القيمة التوفيقية $\hat{Y}_{ij}$.

$$\hat{Y}_{ij} = b_0 + b_1 X_j \tag{4.16}$$

وبالتالي فإن مجموع مربعات الخطأ للنموذج المخفض هو مجموع مربعـات الخطـأ المعتاد SSE:

$$SSE(R) = \sum\sum\left[Y_{ij} - (b_0 + b_1 X_j)\right]^2 \tag{4.17}$$
$$= \sum\sum(Y_{ij} - \hat{Y}_{ij})^2 = SSE$$

وكذلك نعلم أن درجات الحرية المرتبطة مع $SSE(R)$ هي:

$$df_R = n - 2$$

وفي مثال المصرف لدينا من جدول (٤-٥):

$$SSE(R) = SSE = 14{,}741.6$$
$$df_R = 9$$

إحصاءة اختبار

إحصاءة الاختبار الخطّي العام (3.69) هي:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

وهنا تصبح:

$$F^* = \frac{SSE - SSPE}{(n-2) - (n-c)} \div \frac{SSE(F)}{n-c} \tag{4.18}$$

ويسمى الفرق بين مجموعي مربعات الخطأ هنا مجموع مربعات نقص التوفيق ويُشار له بـ $SSLF$:

$$SSLF = SSE - SSPE \tag{4.19}$$

وبالتالي نستطيع التعبير عن إحصاءة الاختبار كما يلي:

$$F^* = \frac{SSLF}{c-2} \div \frac{SSPE}{n-c} = \frac{MSLF}{MSPE} \tag{4.20}$$

حيث يشير $MSLF$ إلى متوسط مربعات نقص التوفيق، ويشير $MSPE$ إلى متوسطات مربعات الخطأ البحت.

نعلم أن قيم F^* الكبيرة تقود إلى استنتاج H_a في الاختبار الخطّي العام. وتصبح قاعدة القرار (3.70) هنا:

$$\begin{aligned} &\text{إذا كانت } F^* \le F(1-\alpha; c-2, n-c) \text{ استنتج } H_0 \\ &\text{وإذا كانت } F^* > F(1-\alpha; c-2, n-c) \text{ استنتج } H_a \end{aligned} \tag{4.21}$$

وفي مثال البنك يمكن حساب إحصاءة الاختبار بسهولة من نتائجنا السابقة:

$SSPE = 1,148.0$

$SSE = 14,741.6$ $\qquad n - c = 11 - 6 = 5$

$SSLF = 14,741.6 - 1,148.0 = 13,593.6$ $\qquad c - 2 = 6 - 2 = 4$

$$F^* = \frac{13,593.6}{4} \div \frac{1,148.0}{5} = \frac{3,398.4}{229.6} = 14.80$$

وإذا اتخذنا مستوى المعنوية 0.01 فنحتاج إلى $F(0.99;4,5) = 11.4$، وحيث إن $F^* = 14.80 > 11.4$، فنستنتج H_a، أي أن دالة الانحدار ليست خطّية. وهذا بالطبع يتطابق مع انطباعنا بالعين المجردة من الشكل (٤–١١). ولبيان القيمة -P لإحصاءة الاختبار، نلاحظ أن

$F^* = 14.80$ تقع بين $F(0.99;4,5) = 11.4$ و $F(0.955;4,5) = 15.6$

وهكذا ينبغي أن تقع القيمة -P بين 0.005 و 0.01 ويمكن تبيان أن القيمة -P هي 0.006.

جدول تحاين

يوضح تعريف مجموع مربعات نقص التوفيق $SSLF$ في (4.19) أننا فككنا مجموع مربعات الخطأ إلى مركّبتين:

$$SSE = SSPE + SSLF \tag{4.22}$$

وهذا التفكيك يتبع من المتطابقة:

$$\underbrace{Y_{ij} - \hat{Y}_{ij}}_{\text{انحراف خطأ}} = \underbrace{Y_{ij} - \overline{Y}_j}_{\text{انحراف خطأ بحت}} + \underbrace{\overline{Y}_j - \hat{Y}_{ij}}_{\text{انحراف نقص توفيق}} \tag{4.23}$$

وتبين هذه المتطابقة أن انحرافات الخطأ في SSE مكونة من مركّبة خطأ بحت ومركّبة نقص توفيق. ويوضّح شكل (٤-١٢) هذا التجزئ للمشاهدة $Y_{13} = 160$ و $Y_3 = 125$ في مثال المصرف.

عندما نربع (4.23) ونجمع فوق كل المشاهدات نحصل على (4.22) لأن مجموع الجداءات يساوي صفرا:

$$\sum\sum(Y_{ij} - \hat{Y}_{ij})^2 = \sum\sum(Y_{ij} - \overline{Y}_j)^2 + \sum\sum(\overline{Y}_j - \hat{Y}_{ij})^2 \tag{4.24}$$

$$SSE \quad = SSPE \quad + SSLF$$

لاحظ من (4.24) أنه يمكننا تعريف مجموع مربعات نقص التوفيق مباشرة كما يلي:

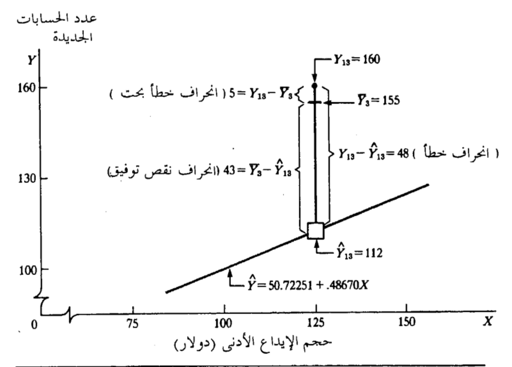
$$SSLF = \sum\sum(\overline{Y}_j - \hat{Y}_{ij})^2 \tag{4.25}$$

وتُفسر العلاقة (4.25) بوضوح لماذا يقيس $SSLF$ نقص التوفيق. فإذا كانت دالة الانحدار الخطّي مناسبة فإن المتوسطات $\overline{Y}_j$ ستكون قريبة من القيمة التوفيقية $\hat{Y}_{ij}$ المحسوبة من دالة الانحدار الخطية المقدَّرة ويكون $SSLF$ صغيرا. ومن جهة أخرى، إذا كانت دالة الانحدارالخطية غير مناسبة فسوف لا تكون المتوسطات $\overline{Y}_j$ قريبة من القيم التوفيقية المحسوبة من دالة الانحدار الخطية المقدَّرة، كما في الشكل (٤-١١) لمثال المصرف، وستكون $SSLF$ كبيرة.

وتشير العلاقة (4.25) أيضا إلى سبب ارتباط $c - 2$ درجة حرية مع $SSLF$، فهناك c متوسطا $\overline{Y}_j$ في مجموع المربعات، وتفقد درجتي حرية، عند تقدير المعلمتين β_0 وβ_1 في دالة

الانحدار الخطّية للحصول على القيم التوفيقية $\hat{Y}_{ij}$.

شكل (٤ ـ١٢) توضيح لتفكيك انحراف الخطأ $(Y_{ij} - \hat{Y}_{ij})$ ـ مثال المصرف.



يمكن إقامة جدول تحاين من تفكيك SSE. ويحوي الجدول (٤ـ٧)أ جدول التحاين العام، متضمنا تفكيك SSE الموضح آنفا ومتوسط المربعات المعني، ويحوي الجدول (٤ـ٧)ب تفكيك التحاين في مثال المصرف.

تعليقات

١ـ كما هو موضح في مثال المصرف، لا حاجة لتوافر مشاهدات متكررة عند كل مستوى من مستويات X كي نطبق الاختبار F لنقص التوفيق. إذ يكفي تكرار المشاهدات عند مستوى واحد فقط أو عند بعض من مستويات X.

٢ـ يمكن إثبات أن توقعي متوسطي المربعات $MSPE$ و$MSLF$ هما كما يلي:

$$E\{MSPE\} = \sigma^2 \quad (4.26)$$

$$E\{MSLF\}=\sigma^2+\frac{\sum n_j\left[\mu_j-(\beta_0+\beta_1X_j)\right]^2}{c-2} \qquad (4.27)$$

وسبب مصطلح "**خطأ بحت**" هو أن $MSPE$ يشكل دائما مقدرا غير منحاز لتباين حد الخطأ σ^2، وذلك أيا كانت دالة الانحدار الحقيقية، والقيمة المتوقعة لـ $MSLF$ هي أيضا σ^2 إذا كانت دالة الانحدار خطّية، ذلك لأن $\mu_j=\beta_0+\beta_1X_j$ عندئذٍ، مما يجعل الحد الثاني في (4.27) صفرا. ومن جهة أخرى، إذا لم تكن دالة الانحدار خطّية، $\mu_j \neq \beta_0+\beta_1X_j$ فإن $E\{MSLF\}$ سيكون أكبر من σ^2. وهكذا فإن قيمة F^* القريبة من 1 تتفق مع دالة انحدار خطّية وتشير قيم F^* الكبيرة إلى أن دالة الانحدار ليست خطّية.

٣ ـ افترض أنه قبل أي تحليل لصلاحية النموذج ترغب في اختبار ما إذا كان $\beta_1=0$ أم لا في مثال المصرف (جدول ٤ـ٥) فستكون إحصاءة الاختبار (3.59):

$$F^*=\frac{MSR}{MSE}=\frac{5,141.3}{1,638.0}=3.14$$

ومن أجل $\alpha=0.10$ و $F(0.90;1,9)=3.36$ نستنتج H_0، أي أن $\beta_1=0$ أو أنه لا توجد صلة خطية بين حجم الإيداع الأدنى (وقيمة الهدية) وبين الحسابات الجديدة. وعلى أي حال، فإن استنتاج عدم وجود علاقة بين هذه المتغيرات قد لا يكون استنتاجا سليما. فمثل هذا الاستقراء يتطلب كون نموذج الانحدار (3.1) مناسبا. وهنا ليس الأمر كذلك، كما رأينا، لأن دالة الانحدار غيرخطّية. وفي الحقيقة توجد علاقة (منحنية) بين حجم الإيداع الأدنى وعدد الحسابات الجديدة، واختبار ما إذا كان $\beta_1=0$ أم لا تحت هذه الظروف له مضامين مختلفة تماما. وهذا يوضح أهمية اختبار صلاحية النموذج دائما قبل القيام بمزيد من الاستقراءات.

٤ ـ يمكن استخدام طريقة الاختبار الخطّي العام الموضحة آنفا لاختبار صلاحية دوال انحدارأخرى وليس فقط الخطّية البسيطة في (4.14). ونحتاج فقط إلى تعديل درجات حرية $SSLF$. وعموما يرتبط مع $SSLF$ و $c-p$ درجة حرية، حيث p عدد المعالم في دالة الانحدار. وفي اختبار دالة انحدار خطّية بسيطة يكون $p=2$ إذ توجد معلمتان β_0 و β_1 في دالة الانحدار.

جدول (٤ ـ٧) جدول تباين لاختبار نقص توفيق دالة الانحدار الخطّية البسيطة وجدول مثال المصرف.

(أ) عام

مصدر الانحراف	SS	df	MS
انحدار	SSR	1	MSR
خطأ	SSE	$n-2$	MSE
نقص توفيق	$SSLF$	$c-2$	$MSLF$
خطأ بحت	$SSPE$	$n-2$	$MSPE$
مجموع	$SSTO$	$n-1$	

(ب) مثال المصرف

مصدر الانحراف	SS	df	MS
انحدار	$SSR = 5,141.3$	1	$MSR = 5,141.3$
خطأ	$SSE = 14,741.6$	9	$MSE = 1,638.0$
نقص توفيق	$SSLF = 13,593.6$	4	$MSLF = 3,398.4$
خطأ بحت	$SSPE = 1,148.0$	5	$MSPE = 229.6$
المجموع	$SSTO = 19,882.9$	10	

٥ ـ يتضمن البديل H_a في (4.14) جميع دوال الانحدار خلاف الخطّية منها. فمثلا تتضمن دالة انحدار تربيعية أو دالة لوغاريتمية. وفي حال استنتاج H_a تكون دراسة الرواسب مفيدة للتعرف على دالة الانحدار المناسبة.

٦ ـ إذا استنتجنا أن النموذج المستخدم H_0 مناسب فإن الممارسة المعتادة هي تفضيل استخدام متوسط مربع الخطأ MSE على متوسط مربع الخطأ البحث $MSPE$ كمقدِّر لـ σ^2، ذلك لأن الأول يتضمن درجات حرية أكثر.

٧ ـ لا تشكل المشاهدات عند المستوى نفسه لـ X تكرارات حقيقية إلا إذا انطوت على محاولات مستقلة بالنسبة لحد الخطأ. فلنفرض في تحليل انحدار للعلاقة بين الصلابة Y وكمية الكربون X في عينات من خليطة معدنية أن حد الخطأ في النموذج يغطي، من بين أشياء أخرى، الأخطاء العشوائية في قياس المحلل للصلابة، وتأثير عوامل إنتاج لا يمكن التحكم فيها، وهي تتغير بصورة عشوائية من عينة إلى أخرى، وتؤثر

على الصلابة. إذا أخذ المحلل قراءتين على صلابة العينة، فسوف لا يشكل ذلك تكرارا أصيلا لأن تأثيرات التغير العشوائي في عوامل الإنتاج تبقى هنا ثابتة لأي عينة بالذات. وللحصول على تكرارات أصيلة ينبغي أن يقيس المحلل عينات مختلفة من الخليطة المعدنية لها المحتوى نفسه من الكربون X بحيث يمكن لجميع التأثيرات التي يغطيها حد الخطأ أن تتغير عشوائيا من مشاهدة مكررة إلى المشاهدة التي تليها.

٨ - من الواضح أن تكرار المشاهدات سيكون أكثر جدوى كلما كنا غير متأكدين من طبيعة دالة الانحدار. وحينما يكون ممكنا، فإنه ينبغي اتخاذ الحيطة للحصول على بعض التكرارات. وإذا لم يكن ممكنا الحصول على تكرارات، فيمكن أحيانا القيام باختبار تقريبي لنقص التوفيق. ولاختبار تقريبي، لا بد من توافر بعض المشاهدات عند مستويات متجاورة لـ X وتكون متوسطات الاستجابة من أجلها قريبة بعضها من بعض. وتُضم هذه المشاهدات المتجاورة إلى بعضها لتشكل ما يشبه التكرارات وذلك لغرض القيام باختبار نقص توفيق.

(٤ - ٦) نظرة إجمالية للتدابير العلاجية

إذا لم يكن نموذج الانحدار الخطّي البسيط (3.1) مناسبا لمجموعة البيانات فإنه يوجد اختباران أساسيان.

١- ترك نموذج الانحدار (3.1) والبحث عن نموذج أكثر صلاحية.

٢- استخدام تحويل للبيانات بحيث يكون نموذج الانحدار (3.1) مناسبا للبيانات بعد تحويلها.

ولكلي الأسلوبين ميزات ومساوىء. فقد يقود الأسلوب الأول إلى نموذج أكثر تعقيدا ويعطي تبصرا أفضل ولكنه ربما قاد أيضا إلى صعوبات في تقدير المعالم. ومن جهة أخرى، فإن استخدام التحويلات بنجاح يقود إلى طرق تقدير بسيطة نسبيا، وقد ينطوي على معالم أقل من النموذج المعقد وهذا مفيد عندما يكون حجم العينة صغيرا.

ومع ذلك فقد تحجب التحويلات الروابط الداخلية الأساسية بين المتغيرات علما أنها يمكن، في أحيان أخرى، أن تزيدها وضوحا.

وسوف ندرس استخدام التحويلات خلال هذا الفصل واستخدام نماذج أكثر تعقيدا في فصول قادمة. ونقدّم أولا مراجعة مختصرة للتدابير العلاجية.

عدم خطية دالة الانحدار

إذا لم تكن دالة الانحدار خطّية فإن الأسلوب المباشر هو تعديل نموذج الانحدار (3.1) من حيث طبيعة دالة الانحدار. فمثلا يمكن استخدام دالة انحدار تربيعية:

$$E\{Y\} = \beta_0 + \beta_1 X + \beta_2 X^2$$

أو دالة انحدار رأسية:

$$E\{Y\} = \beta_0 \beta_1^X$$

وفي الفصل التاسع، نناقش نماذج تكون دالة الانحدار فيها كثيرة حدود.

ويستخدم أسلوب التحويل تحويلا يؤمّن، بصورة تقريبية على الأقل، خطّية دالة انحدار غير خطّية. ونناقش في الفقرة القادمة استخدام تحويلات تجعل دوال الانحدار خطّية.

ثبات تباين الخطأ

إذا كان تباين الخطأ غير ثابت وإنما يتغير بصورة نمطية، فإن الأسلوب المباشر هو تعديل النموذج بحيث يسمح بمثل هذا التغير ثم استخدام طريقة المربعات الدنيا المرجحة للحصول على مقدّرات للمعالم. ونناقش في الفصل الحادي عشر استخدام المربعات الدنيا المرجحة لهذا الغرض.

ويمكن أن تكون التحويلات فعّالة أيضا في جعل التباين مستقرا. ونناقش بعضا من هذه التحويلات في الفقرة التالية.

عدم استقلالية حدود الخطأ

إذا كانت حدود الخطأ مرتبطة فإن التدبير العلاجي المباشر هو العمل مع نموذج يسمح بحدود خطأ مرتبطة. ونناقش مثل هذا النموذج في الفصل الثالث عشر. وغالبا

مايكون التحويل العلاجي البسيط الذي يتعامل مع الفروقات الأولى تحويلا مفيدا وسنناقش هذا الموضوع أيضا في الفصل الثالث عشر.

عدم طبيعية حدود الخطأ

كثيرا ما يتماشى نقص الطبيعية مع عدم ثبات تباينات الخطأ يدا بيد. ولحسن الحظ، فإنه غالبا مانواجه الحالة التي يكون فيها التحويل نفسه مفيدا في جعل التباين مستقرا ومفيدا في جعل حدود الخطأ طبيعية. ولذلك فمن المرغوب فيه الإفادة أولا من التحويل الذي يجعل تباين الخطأ مستقرا ثم تُدرس الرواسب لرؤية ما إذا كانت لاتزال توجد انحرافات جدّية عن الطبيعية. ونناقش تحويلات للوصول إلى الطبيعية في الفقرة القادمة.

حذف متغيرات مستقلة مهمة

عندما يشير تحليل الراسب إلى أن متغيرا مستقلا مهما قد حُذف من النموذج، يكون الحل في تعديل النموذج. وفي الفصل السابع ومابعده من فصول، نناقش تحليل الانحدار المتعدّد وفيه يستخدم متغيران مستقلان أو أكثر.

(٤-٧) تحويلات

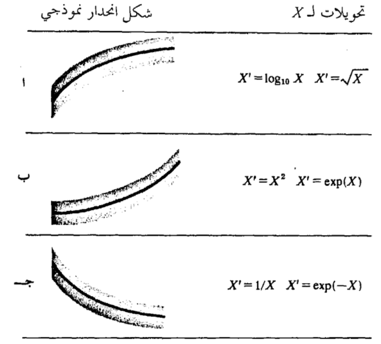
نعتبر الآن بتفصيل أكثر استخدام التحويلات لأحد المتغيرين الأصليين أو كليهما قبل تنفيذ تحليل الانحدار. وغالبا ماتكون تحويلات بسيطة للمتغير التابع Y أو للمتغير المستقل X أو كليهما كافية لجعل نموذج الانحدار الخطّي البسيط مناسبا للبيانات بعد تحويلها.

تحويلات لعلاقات غير خطّية فقط

نعتبر أولا تحويلات إلى الخطية لعلاقات انحدار غير خطّية وذلك عندما يكون توزيع حدود الأخطاء قريبا من التوزيع الطبيعي، ولحدود الخطأ تباين ثابت تقريبا. وفي هذه الحالة ينبغي تجربة تحويلات على X. وسبب عدم تفضيل التحويلات على Y هو أن تحويلا على Y مثل $Y=\sqrt{Y}$ قد يغيّر كثيرا من شكل توزيع حدود الخطأ مبتعدا عن التوزيع الطبيعي وربما قاد أيضا إلى تباينات مختلفة جوهريا لحد الخطأ.

ويحوي الشكل (٤-١٣) بعض علاقات انحدار غير خطّية نموذجية بتباين حد خطأ ثابت، كما يقدم بعض التحويلات البسيطة على X التي قد تفيد في جعل علاقة الانحدار خطّية دون التأثير على توزيعات Y. ويمكن تجربة تحويلات بديلة. وعندئذ ينبغي القيام برسومات نقطية ورسومات رواسب مرتكزة على كل من هذه التحويلات ثم تحليلها لتقرير أيها أكثر فعالية.

شكل (٤-١٣) أشكال نموذجية لانحدار غير خطّي بتباين خطأ ثابت وتحويلات بسيطة لـ X.



مثال. يعرض العمودان ١ و ٢ من الجدول (٤-٨) بيانات تجريبية لمتدربي مبيعات، وهي تتضمن عدد أيام التدريب التي تلقاها المتدرب X، ودرجة الأداء Y، في مجموعة من حالات البيع المحسوبة على سبيل التحاكي. وهذه المشاهدات موضحة في رسم انتشار في الشكل (٤-١٤)أ. ومن الواضح أن علاقة الانحدار تبدو منحنية، ولذلك لايبدو أن دالة الانحدار الخطّية البسيطة (3.1) مناسبة. وحيث إن التشتيت عند مستويات X المختلفة يبدو ثابتا تقريبا، فسنعتبر تحويلا على X. وبالاستناد إلى الرسم

النموذجي في شكل (٤-١٣)أ سنعتبر مبدئيا تحويل الجذر التربيعي وعُرضت القيم المحولة في العمود $X'=\sqrt{X}$ من الجدول (٤-٨).

وفي شكل (٤-١٤)ب رُسمت البيانات نفسها بعد تحويل المتغير المستقل إلى $X'=\sqrt{X}$. ولاحظ الآن أن رسم الانتشار يوضح بصورة معقولة وجود علاقة خطية.

جدول (٤-٨) حسابات انحدار مع تحويل جذر تربيعي لـX، مثال تدريب مبيعات

مندوبو مبيعات	(١) أيام التدريب	(٢) درجة الأداء	(٣) $X_i'=\sqrt{X_i}$	(٤) $X_i'=Y_i$	(٥) $(X_i')^2$
1	0.5	46	0.70711	32.527	0.5
2	0.5	51	0.70711	36.062	0.5
3	1.0	71	1.00000	71.000	1.0
4	1.0	75	1.00000	75.000	1.0
5	1.5	92	1.22474	112.677	1.5
6	1.5	99	1.22474	121.250	1.5
7	2.0	105	1.41421	148.492	2.0
8	2.0	112	1.41421	158.392	2.0
9	2.5	121	1.58114	191.318	2.5
10	2.5	125	1.58114	197.643	2.5
المجموع	15.0	897	11.85440	1,144.361	15.0

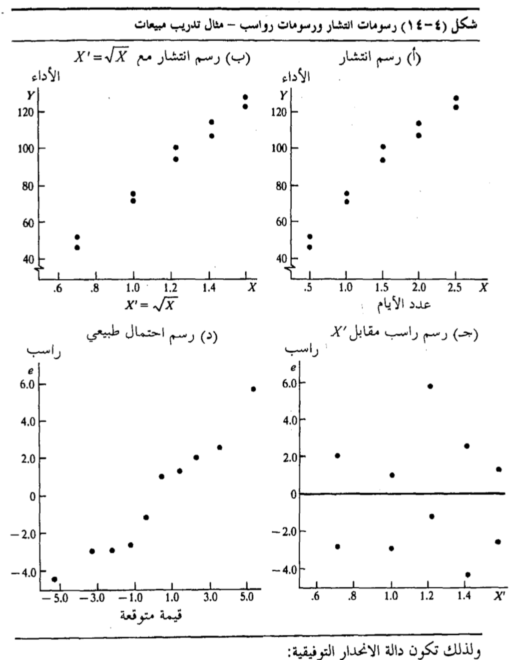
والتشتت في مخطط الانتشار عند مستويات X المختلفة هو نفسه كما سبق، حيث إننا لم نطبّق تحويلا على Y.

ولمزيد من التحقيق عما إذا كان نموذج الانحدار البسيط (3.1) مناسبا الآن، نقوم بتوفيقه لبيانات X المحوّلة. وتتم حسابات الانحدار مع بيانات X المحوّلة بالطريقة المعتادة. ويحوي الجدول (٤-٨) حسابات المربعات الدنيا الضرورية. وحيث إن X' تلعب الآن دور X في جميع العلاقات السابقة، فإننا نجد:

$$b_1=\frac{\sum X_i' Y_i-\frac{\sum X_i'\sum Y_i}{n}}{\sum(X_i')^2-\frac{(\sum X_i')^2}{n}}=\frac{1{,}144.361-\frac{11.85440(897)}{10}}{15.0-\frac{(11.85440)^2}{10}}=85.5259$$

$$b_0=\frac{1}{n}\left(\sum Y_i-b_1\sum X_i'\right)=\frac{1}{10}\left[897-85.5259(11.85440)\right]=-11.6858$$

شكل (٤-١٤) رسومات انتشار ورسومات رواسب - مثال تدريب مبيعات



ولذلك تكون دالة الانحدار التوفيقية:

$$\hat{Y}=-11.69+85.53X'$$

ويحوي الشكل (٤-١٤)جـ رسم الرواسب مقابل X'. ولايوجد دليل على نقص التوفيق أو على اختلاف قوي في تباينات الخطأ. ويحوي الشكل (٤-١٤)د رسم

احتمال طبيعي للرواسب. ولا توجد مؤشرات قوية على انحرافات عن الطبيعية موضحة في هذا الرسم. وتعزِّز هذا الاستنتاج القيمة المرتفعة لمعامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية، 0.972 (انظر الجدول ٤-٣). وهكذا يبدو أن نموذج الانحراف الخطّي البسيط (3.1) هو النموذج المناسب هنا للبيانات المحوَّلة. ويمكن الحصول على دالة الانحدار التوفيقية بالوحدات الأصلية لِـ X إذا رغبنا:

$$\hat{Y} = -11.69 + 85.53\sqrt{X}$$

ملاحظة

قد يكون مفيدا، أحيانا، إدخال ثابت في عملية التحويل. فمثلا، إذا كانت بعض بيانات X قريبة من الصفر ونرغب في استخدام تحويل المقلوب، فيمكن إزاحة نقطة الأصل واستخدام التحويل $X' = 1/(X+k)$ حيث k ثابت نختاره بصورة مناسبة.

تحويلات لمعالجة اللاطبيعية وعدم تساوي تباينات الخطأ

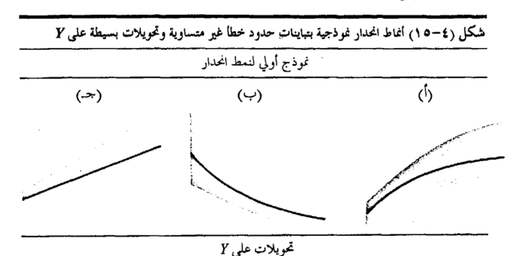
كثيرا ما تُظهر مشكلتنا عدم تساوي تباينات الخطأ ولاطبيعية حدود الخطأ معا. ولعلاج هذا الانحراف عن نموذج الانحدار الخطي البسيط (3.1)، نحتاج إلى تحويل Y، فأشكال وانتشارات توزيعات Y في حاجة إلى تغيير. وقد يساعد مثل هذا التحويل في الوقت نفسه على جعل علاقة الانحدار المنحنية خطّية. وفي أحيان أخرى، قد نحتاج في الوقت نفسه إلى تحويل X أيضا للحصول على علاقة انحدار خطّية أو الإبقاء عليها.

وكثيرا ما يتخذ انحراف اللاطبيعية وعدم تساوي التباينات عن نموذج الإنحدار (3.1) شكل التواء، وتشتت في توزيعات حدود الخطأ تتزايد مع ازدياد متوسط الاستجابة $E\{Y\}$. فمثلا، في انحدار مصاريف الأسرة السنوية في السياحة Y على دخلها السنوي X هناك اتجاه لتشتت، والالتواء موجب (أي وجود مصروفات سنوية مرتفعة جدا) أكبر في حالة الأسر ذات الدخول المرتفعة منهما في حالة الأسر ذات الدخول المنخفضة التي تنحو باستمرار إلى أن تكون نفقاتها أقل كثيرا. ويحوي الشكل (٤-١٥) بعض علاقات انحدار نموذجية حيث يزداد الالتواء وتباين الخطأ مع متوسط الاستجابة $E\{Y\}$. ويعرض هذا الشكل أيضا بعض التحويلات البسيطة على Y مما قد

يكون مفيدا لهذه الحالات. ويمكن تجربة عدة تحويلات بديلة على Y، إضافة إلى بعض التحويلات على X في الوقت نفسه، وينبغي إعداد رسومات انتشار ورسومات راسب لتحديد التحويل الأكثر فعالية.

شكل (٤-١٥) أنماط انحدار نموذجية بتباينات حدود خطأ غير متساوية وتحويلات بسيطة على Y

نموذج أولي لنمط انحدار



تحويلات على Y

$$Y' = \sqrt{Y}$$
$$Y' = \log_{10} Y$$
$$Y' = 1/Y$$

ملاحظة: قد تكون التحويلات على X في الوقت نفسه مفيدة.

مثال. في العمودين ١ و ٢ من الجدول (٤-٩) نقدّم بيانات عن العمر X ومستوى البلازما بوليامين Y (*Polyamine*) وذلك لـ 25 من الأطفال الأصحّاء. ورُسمت هذه البيانات في الشكل (٤-١٦)أ كرسم انتشار ونلاحظ علاقة منحنية واضحة، بالاضافة إلى تشتت أكبر في الأطفال الصغار منه في الأطفال الأكبر.

وبالاستناد إلى أنماط الانحدار في الشكل (٤-١٥)ب، سنستخدم أولا التحويل اللوغاريتمي $Y' = Log_{10}Y$. وقد عُرضت قيم Y بعد التحويل في العمود (٣) من الجدول (٤-٩). ويحوي الشكل (٤-١٦)ب رسم انتشار للبيانات بعد التحويل. ونلاحظ أن التحويل لم يؤدِ فقط إلى انحدار خطّي معقول ولكن التشتت عند مستويات X المختلفة أصبح أيضا ثابتا بصورة مقبولة.

ولمزيد من التحقق من معقولية التحويل $Y' = Log_{10}Y$، قمنا بتوفيق نموذج الانحدار الخطّي البسيط (3.1) لبيانات Y بعد التحويل وحصلنا على:

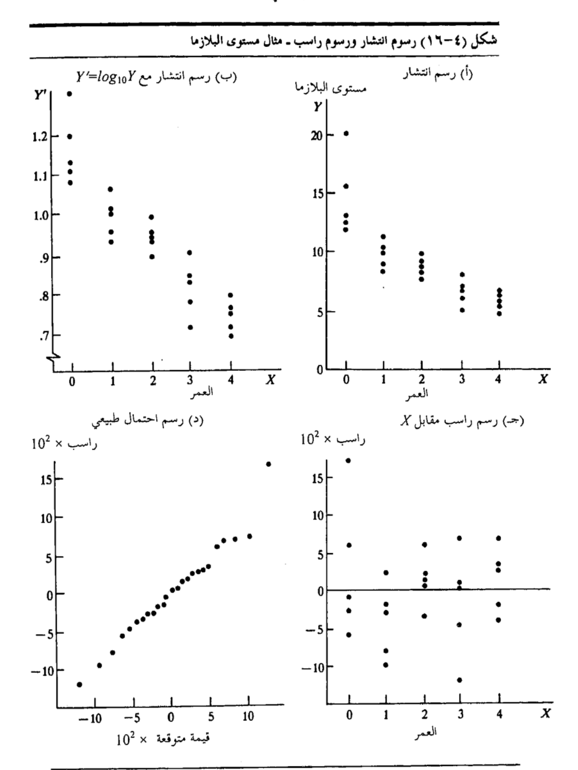
$$\hat{Y}' = 1.135 - 0.1023X$$

جدول (٤–٩) بيانات مثال مستوى البلازما والتحويل اللوغاريتمي لـ Y

الطفل i	(١) العمر X_i	(٢) مستوى البلازما Y_i	(٣) $Y_i' = \log_{10} Y_i$
1	(حديث الولادة) 0	13.44	1.1284
2	(حديث الولادة) 0	12.84	1.1086
3	(حديث الولادة) 0	11.91	1.0759
4	(حديث الولادة) 0	20.09	1.3030
5	(حديث الولادة) 0	15.60	1.1931
6	1.0	10.11	1.0048
7	1.0	11.38	1.0561
8	1.0	10.28	1.0120
9	1.0	8.96	0.9523
10	1.0	8.59	0.9340
11	2.0	9.83	0.9926
12	2.0	9.00	0.9542
13	2.0	8.65	0.9370
14	2.0	7.85	0.8949
15	2.0	8.88	0.9484
16	3.0	7.94	0.8998
17	3.0	6.01	0.7789
18	3.0	5.14	0.7110
19	3.0	6.90	0.8388
20	3.0	6.77	0.8306
21	4.0	4.86	0.6866
22	4.0	5.10	0.7076
23	4.0	5.67	0.7536
24	4.0	5.75	0.7597
25	4.0	6.23	0.7945

لم تعطَ حسابات الانحدار لأنها،دائما، الحسابات نفسها باستثناء أن Y' يحل محل Y في العلاقة (2.10). وعُرض رسم الرواسب مقابل X في الشكل (٤–١٦)جـ، كما عُرض رسم احتمال طبيعي للرواسب في الشكل (٤–١٦)د. ومعامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية هو 0.981 وجميع هذه الأدلة تدعم صلاحية نموذج الانحدار (3.1) لبيانات Y بعد التحويل.

شكل (٤-١٦) رسوم انتشار ورسوم راسب ـ مثال مستوى البلازما



وإذا رغبنا في عرض دالة الانحدار المقدرة بالوحدات الأصلية لـ Y نـأخذ ببساطة اللوغاريتم المعاكس لـ $\hat{Y}'$ لنجد:

$$\hat{Y} = anti\log_{10}(1.135 - 0.1023X)$$

ملاحظة

قد نرغب أحيانا إدخال ثابت في تحويل Y، كالحالة التي تكون Y فيها سالبة. فمثلا، يكون التحويل اللوغاريتمي الذي ينقل نقطة الأصل بالنسبة لـ Y ليجعل كل المشاهدات Y موجبة هو $Y' = log_{10}(Y + k)$ حيث k ثابت نختاره بصورة مناسبة.

تحويلات بوكس – كوكس (Box - Cox)

طور بوكس وكوكس (مرجع 4.4) طريقة لاختيار تحويل من عائلة تحويلات قوى لـ Y وهذه الطريقة مفيدة لتصحيح التواء توزيعات حدود الخطأ وعدم تساوي تباينات الخطأ، وعدم خطّية دالة الانحدار. وعائلة تحويلات القوى هي من الشكل:

$$Y' = Y^{\lambda} \tag{4.28}$$

حيث λ معلمة تُحدَّد من البيانات. لاحظ أن هذه العائلة تشمل التحويلات البسيطة التالية:

$\lambda = 2$	$Y' = Y^2$
$\lambda = 0.5$	$Y' = \sqrt{Y}$
$\lambda = 0$	$Y' = \log_e Y$
$\lambda = -0.5$	$Y' = \frac{1}{\sqrt{Y}}$
$\lambda = -1.0$	$Y' = \frac{1}{Y}$

والمعيار في تحديد قيمة المعلمة λ المناسبة لتحويل Y في طريقة بوكس ـ كوكس هي إيجاد قيمة Y التي تجعل مجموع مربعات الخطأ SSE لانحدار خطّي يستند إلى ذلك التحويل أصغر مايمكن. وتتوافر برامج حاسب آلي لإيجاد قيمة λ المناسبة. وكبديل، يمكن اختيار عدد من قيم λ والقيام بالتحويل المقابل لكل منها ثم توفيق دالة الانحدار الخطّية لبيانات Y بعد التحويل وحساب SSE لكل توفيق. ومن ثم اختيار القيمة التي تجعل SSE أصغر ما يمكن.

وإذا رغبنا، يمكن إجراء بحث أدق في جوار λ عن القيمة التي تجعل SSE أصغر ما يمكن. إلا أن طريقة بوكس ـ كوكس تُستخدم عادة لتكون فقط مرشدا في اختيار تحويل، مما لايترك حاجة للقيم الدقيقة جدا لِ λ، وعلى أي حال ينبغي الإفادة من رسمي الانتشار والراسب للتحقق من صلاحية التحويل الذي تحدده طريقة بوكس ـ كوكس.

وبما أن قوة التحويل تؤثر في حجم مجموع مربعات الخطأ SSE، فإننا نحتاج إما إلى تعديل لـ SSE يأخذ هذا في الاعتبار أو إلى استخدام متغير معياري يترك SSE غير متأثر بقيمة λ. ويؤخذ عادة بالأسلوب الأخير ويستخدم المتغير المعياري التالي:

$$W = \begin{cases} K_1(Y^{\lambda} - 1) & \lambda \neq 0 \\ K_2(\log_e Y) & \lambda = 0 \end{cases} \quad (4.29)$$

حيث:

$$K_2 = \left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}} \quad (4.29a)$$

$$K_1 = \frac{1}{\lambda K_2^{\lambda-1}} \quad (4.29b)$$

لاحظ أن K_2 هي المتوسط الهندسي لمشاهدات Y.

مثال. يحوي الجدول (٤-١٠) نتائج بوكس ـ كوكس لمثال مستويات البلازما. اختيرت قيم لـ λ تتراوح بين 1.0- إلى 1.0، ثم استكملت التحويلات وتمَّ توفيق انحدار خطّي لكل منها. فمثلا، من أجل $\lambda = 0.5$، أخذ التحويل $W = K_1(\sqrt{Y} - 1)$ وجرى توفيق الانحدار الخطّي لِ W على X. ولدينا $SSE = 48.4$ من أجل هذا التوفيق لانحدار خطّي.

ونلاحظ من الجدول (٤-١٠) أن طريقة بوكس ـ كوكس تحدّد أساسا قريبا من $\lambda = -0.50$، كقيمة مناسبة. وعلى أي حال، فإن SSE، كدالة في λ، مستقرة بصورة مقبولة ضمن المدى من قرب الصفر إلى 1.0- وبالتالي فإن SSE الاختيار السابق لتحويل لوغاريتمي $Y' = log_{10}Y$. مستويات البلازما ليس تحويلا غير منطقي وفقا لطريقة بوكس ـ كوكس. وأحد أسباب اختيار التحويل اللوغاريتمي هنا هو سهولة تفسيره. واستخدام لوغاريتم للأساس 10 بدلا من اللوغاريتم الطبيعي لايؤثر، بالطبع، في صلاحية التحويل اللوغاريتمي.

جدول (٤-١٠) نتائج بوكس - كوكس لمثال مستويات البلازما

λ	*SSE*
1.0	78.0
0.9	70.4
0.7	57.8
0.5	48.4
0.3	41.4
0.1	36.4
0	34.5
−0.1	33.1
−0.3	31.2
−0.4	30.7
−0.5	30.6
−0.6	30.7
−0.7	31.1
−0.9	32.7
−1.0	33.9

تعليقات

١- يمكن الاستفادة، أحيانا، من اعتبارات مسبقة أو نظرية للمساعدة في اختيار تحويل مناسب. فمثلا، عندما يكون شكل الانتشار في دراسة العلاقة بين سعر سلعة X وكمية الطلب Y كما في الشكل (٤-١٥)ب، فقد يفضل الاقتصاديون تحويلا لوغاريتميا لكل من Y و X لأن ميل خط الانحدار الخاص بالمتغيرين بعد التحويل يقيس عندئذ مرونة الطلب السعرية. والتفسير الشائع للميل عندئذ هو أنه يبين النسبة المئوية للتغير في كمية الطلب لكل 1 بالمائة تغير في السعر، ومن المعروف أن التغيرين هما في اتجاهين متعاكسين.

وبصورة مشابهة، قد يفضل العلماء تحويلين لوغاريتميين لكل من Y و X عند دراسة العلاقة بين تناقص النشاط الإشعاعي Y لمادة والزمن X وذلك من أجل علاقة منحنية من النوع الموضّح في الشكل (٤-١٥)ب لأن ميل خط الانحدار للمتغيرين بعد التحويل يقيس عندئذ معدّل التناقص.

٢- بعد اختيار تحويل، مبدئيا، ينبغي القيام برسوم رواسب وبقية التحليلات الموصوفة سابقا للتحقق من أن نموذج الانحدار الخطّي البسيط (3.1) مناسب للبيانات بعد التحويل.

٣- عندما تُستخدم النماذج بعد التحويل فإن للمقدّرات b_0 و b_1 التي نحصل عليها من المربعات الدنيا خواص المربعات الدنيا بالنسبة للمشاهدات بعد التحويل وليس للمشاهدات الأصلية.

المراجع

[4.1] Dixon, W. J., (chief editor). *BMDP Statisitcal Software Manual, vols. 1 and* 2. Berkeley, Calif.: University of Californai Press, 1988.

[4.2] MINITAB. *Reference Manual, Releas* 7. State College, Pa.: Minitab, Inc., 1989.

[4.3] Barnett, V. and Lewis, T., *Outliers in Statistical Data, 2nd ed.* New York: John Wiley & Sons, 1984.

[4.4] Box, G. E. P. and Cox, D. R."An Analysis of Transformations." *Journal of the Royal Statistical Society B* 26 (1964), 211 - 243.

مسائل

(٤-١) ميّز بين: (١) راسب وراسب معياري، (٢) $E(\varepsilon_i)=0$ و $\bar{e} = 0$ (٣) حد خطأ وراسب.

(٤-٢) قم بإعداد رسم راسب نموذجي لكل من الحالات التالية: (١) تباين خطأ متناقص مع X و (٢) الشكل الصحيح لدالة انحدار هو الشكل U، ولكن قمنا بتوفيق دالة انحدار خطّية.

(٤-٣) بالعودة إلى مسألة **المعدل التراكمي** (٢-١٧) كانت القيم التوفيقية والرواسب كما يلي:

i	1	2	3	4	5	6	7	8	9	10
$\hat{Y}_i$	2.92	2.33	2.25	1.58	2.08	3.51	3.34	2.67	2.25	1.91
e_i	0.18	-0.03	0.75	0.32	0.42	0.19	0.06	-0.07	0.55	-0.31

i	11	12	13	14	15	16	17	18	19	20
$\hat{Y}_i$	2.25	1.74	3.26	2.50	1.91	2.16	3.59	2.50	2.84	2.42
e_i	-.75	.46	.54	-.50	-.51	-.39	-.39	-.20	.06	-.42

ا - قم بإعداد رسم صندوقي لدرجات اختبار الدخول X_i. هل هناك أي ميزة جديرة بالملاحظة في هذا الرسم؟

ب - قم بإعداد رسم نقطي للرواسب. ماهي المعلومات التي يقدّمها هذا الرسم؟

جـ - ارسم الرواسب e_i مقابل القيم التوفيقية $\hat{Y}_i$. ماهي الانحرافات عن نموذج الانحدار (3.1) التي يمكن دراستها من هذا الرسم؟ وماذا اكتشفت؟.

د - قم بإعداد رسم احتمال طبيعي للرواسب، أوجد كذلك معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية. اختبر معقولية فرضية الطبيعية هنا مستخدما جدول (٤-٣) و$\alpha = 0.05$. ماذا تستنتج؟

هـ - فيما يلي معلومات عن كل طالب حول متغيرين لم يشملهما النموذج، وهما، درجة اختبار ذكاء X_2 ومعدل الثانوية X_3. ارسم الرواسب مقابل X_2 و X_3 في بيانين منفصلين للتحقق مما إذا كان يمكن تحسين النموذج بإدخال أي من هذين المتغيرين. ماذا تستنتج؟

10	9	8	7	6	5	4	3	2	1	i
111	117	121	115	125	110	107	118	113	105	X_2
2.9	3.1	3.1	3.5	2.4	3.0	2.4	3.1	2.8	2.9	X_3

20	19	18	17	16	15	14	13	12	11	i
108	116	109	119	110	122	132	120	114	123	X_2
2.7	2.6	3.4	3.3	2.8	3.0	2.6	3.4	3.3	3.2	X_3

(٤-٤) عُد إلى مسألة **صيانة الحاسبات اليدوية** (٢-١٨)، القيم التوفيقية والرواسب هي:

9	8	7	6	5	4	3	2	1	i
56.5	41.9	100.8	56.6	71.4	12.4	71.4	86.1	100.8	X_2
-3.6	-2.9	.2	5.4	3.6	-2.4	6.6	-.1	-3.8	X_3

18	17	16	15	14	13	12	11	10	i
71.4	56.6	12.4	100.8	71.4	27.2	71.4	115.6	27.2	X_2
-7.6	-7.6	4.6	4.2	-.4	-2.2	-6.4	2.4	5.8	X_3

ا - قم بإعداد رسم نقطي لعدد الآلات المصانة X_i. ماهي المعلومات التي يقدّمها هذا الرسم ؟ هل هناك أية مشاهدات قاصية بالنسبة لهذا المتغير؟

ب ـ المشاهدات معطاة وفق ترتيبها الزمني. قم بإعداد رسم زمني لعدد الآلات المصانة ؟ ماذا يوضّح رسمك؟

جـ ـ قم بإعداد رسم جذع وورقة للرواسب. هل هناك أية ميزات جديرة بالملاحظة في هذا الرسم؟

د ـ قم بإعداد رسم رواسب لـ e_i مقابل $\hat{Y}_i$ و e_i مقابل X_i، في بيانين منفصلين. هل يقدّم هذان الرسمان المعلومات نفسها؟ ماهي الانحرافات عن نموذج الانحدار (3.1) التي يمكن دراستها من هذين الرسمين؟ اعرض مرئياتك.

هـ ـ قم بإعداد رسم احتمال طبيعي للرواسب. وأوجد كذلك معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية. هل تبدو فرضية الطبيعية مقبولة هنا؟ استخدم جدول (٤-٣) و$\alpha = 0.10$.

و ـ قم بإعداد رسم زمني للرواسب للتحقق مما إذا كانت حدود الخطأ مرتبطة فوق الزمن. ماهو استنتاجك؟.

ز ـ فيما يلي معلومات عن متغيرين لم يشملهما نموذج الانحدار، وهما، متوسط عمر التشغيل للآلات المصانة عند نداء خدمة (X_2، بالأشهر) وسنوات الخبرة لرجل الصيانة الذي يجيب النداء (X_3). ارسم الرواسب مقابل X_2 و X_3 في بيانين منفصلين للتحقق مما إذا كان يمكن تحسين النموذج بإدخال أي من هذين المتغيرين أو كليهما. ماذا تستنتج؟

i	1	2	3	4	5	6	7	8	9
X_2	12	21	38	16	25	32	18	14	12
X_3	3	6	2	2	3	4	5	2	3

i	10	11	12	13	14	15	16	17	18
X_2	35	20	8	15	17	28	29	9	14
X_3	6	5	3	5	6	3	5	3	6

(٤-٥) بالعودة إلى مسألة **تكسر الشحنات** (٢-١٩)

ا ـ قم بإعداد رسم نقطي لعدد التحويلات X_i. هل يبدو توزيع عدد التحويلات غير متماثل؟

ب ـ المشـاهدات معطـاة وفـق ترتيبهـا الزمـني. قـم بـإعداد رسـم زمـني لعـدد التحويلات. هل تظهر أية نمطية نظامية في رسمك ؟ ناقش.

جـ ـ أوجد الرواسب e_i وقم بـإعداد رسـم جـذع وورقـة للرواسـب ؟ مـاهي المعلومات التي يقدّمها رسمك؟.

د ـ ارسم الرواسب e_i مقابل X_i للتحقق مما إذا كانت تتضح أية انحرافات عن نموذج الانحدار (3.1). ماهو استنتاجك؟.

هـ ـ قم بإعداد رسم احتمال طبيعي للرواسب. وأوجد كذلك معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية للتحقق مما إذا كـانت فرضية الطبيعية معقولة هنا. استخدم جدول (٤-٣) ومسـتوى معنويـة 0.01 ماذا تستنتج؟

و ـ قم بإعداد رسم زمني للرواسب. ماهي المعلومات التي يقدمها رسمك؟

(٤-٦) بالعودة إلى مسألة **صلابة البلاستيك** (٢-٢٠)

ا ـ أوجد الرواسب e_i وقم بإعداد رسم صندوقي للرواسب. ماهي المعلومـات التي يقدّمها رسمك ؟.

ب ـ ارسم الرواسب e_i مقابل القيم التوفيقية $\hat{Y}_i$ للتحقق مما إذا كـانت هنـاك انحرافات واضحة عن نموذج الانحدار (3.1)، اعرض مرئياتك.

جـ ـ قم بإعداد رسم احتمال طبيعي للرواسب. وأوجد كذلك معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية. هل تبدو فرضيـة الطبيعيـة معقولة هنا؟ استخدم جدول (٤-٣) و$\alpha = 0.05$.

د ـ أوجد الرواسب المعيارية. قـارن التكـرارات الحقيقيـة مقـابل التكـرارات المتوقعة تحت الطبيعية، مستخدما المئينات 25، 50 و75 لتوزيع t الموافـق. هل تنسجم المعلومات التي تقدمها هذه المقارنات مع ما وجدته من رسم الاحتمال الطبيعي في الجزء (جـ).

(٤-٧) عُد إلى مسألة **كتلة العضلة** (٢-٢٥)

ا - قم بإعداد رسم جذع وورقة للأعمار X_i هل ينسجم هذا الرسم مع الاختيار العشوائي للنساء من كل شريحة عمرية بعشر سنوات.

ب - أوجد الرواسب e_i وقم بإعداد رسم تكرار نقطي للرواسب. ماذا يوضّح رسمك؟.

جـ - ارسم الرواسب e_i مقابل $\hat{Y}_i$ ، وكذلك مقابل X_i في بيانين منفصلين للتحقق مما إذا كان هناك أي انحراف واضح عن نموذج الانحدار (3.1). هل يقدم الرسمان المعلومات نفسها ؟.

د - قم بإعداد رسم احتمال طبيعي للرواسب. وأوجد كذلك معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية للتحقق مما إذا كان فرض الطبيعية معقولا هنا، استخدم جدول (٤-٣) ومستوى معنوية 0.10 ماذا تستنتج؟

هـ - أوجد الرواسب المعيارية. قارن التكرارات الحقيقية مقابل التكرارات المتوقعة تحت الطبيعية مستخدما المئينات 25، 75 و 90 من توزيع t الموافق. هل تنسجم المعلومات التي تقدمها هذه المقارنات مع ما وجدته من رسم الاحتمال الطبيعي في الجزء (د)؟.

(٤-٨) بالعودة إلى مسألة **معدل السرقة** (٢-٢٦)

ا - قم بإعداد رسم جذع وورقة للكثافة السكانية في المدينة X ما هي المعلومات التي يقدّمها رسمك ؟

ب - أوجد الرواسب e_i وقم بإعداد رسم صندوقي للرواسب. هل يبدو توزيع الرواسب متماثلا ؟.

جـ - قم بإعداد رسم رواسب لـ e_i مقابل $\hat{Y}_i$. ماذا يوضح الرسم ؟

د - قم بإعداد رسم احتمال طبيعي للرواسب وأوجد كذلك معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية. اختبر معقولية فرضية الطبيعية مستخدما جدول (٤-٣) و$\alpha = 0.05$ ماهو استنتاجك؟

هـ ـ أوجد الرواسب المعيارية. قارن التكرارات الحقيقية مقابل قيمها المتوقعة تحت الطبيعية مستخدما المئينات الخمسين، الخامس والسبعين، والتسعين من توزيع t الموافق. هل تنسجم المعلومات الناتجة من هذه المقارنات مع ما وجدته من رسم الاحتمال الطبيعي في الجزء (د)؟.

(٤-٩) **استهلاك الكهرباء**. استخدم اقتصادي يدرس العلاقة بين استهلاك أسرة للكهرباء Y وعدد الغرف في المنزل X، نموذج الانحدار الخطّي (3.1) وحصل على الرواسب التالية:

10	9	8	7	6	5	4	3	2	1	i
11	10	9	7	6	6	5	4	3	2	X_i
.5	.7	-.9	-1.2	-1.2	-2.3	-2.0	-1.7	2.9	3.2	e_i

ارسم الرواسب e_i مقابل X_i ما هي المشكلة التي يبدو أنها موجودة هنا؟ هل يخفف تحويل للبيانات هذه المشكلة؟

(٤-١٠) **الدخل الفردي**. استخدم اختصاصي اجتماعي نموذج الانحدار الخطّي (3.1) لربط الدخل الإجمالي Y بمعدّل عدد سنوات الدراسة X في 12 مدينة وكانت القيم التوفيقية $\hat{Y}_i$ والرواسب المعيارية $e_i/\sqrt{MSE}$ كالتالي:

6	5	4	3	2	1	i
12.4	10.2	9.6	10.2	9.3	9.9	$\hat{Y}_i$
-.17	.65	.43	-.76	.81	-1.12	$e_i/\sqrt{MSE}$

12	11	10	9	8	7	i
13.1	11.2	15.6	9.2	9.6	14.3	X_2
.32	.74	-3.78	-.53	1.79	1.62	X_3

ا ـ ارسم الرواسب المعيارية مقابل القيم التوفيقية، ماذا يقترح الرسم؟

ب ـ كم عدد الرواسب المعيارية الواقعة خارج الفترة (1,1-)؟ بالتقريب كم العدد الذي تتوقع أن تراه إذا كان النموذج ملائما ؟.

(٤-١١) **تركيز الدواء**. يستخدم صيدلي نموذج الانحدار الخطي (3.1) لدراسة العلاقة بين تركيز الدواء في البلازما (Y) ولوغاريتم جرعة الدواء (X) وكانت الرواسب ومستويات لوغاريتم الجرعة كما يلي:

9	8	7	6	5	4	3	2	1	i
1	0	-1	1	0	-1	1	0	-1	X_i
-4.5	2.6	-.6	4.2	-1.7	.3	-3.4	2.1	.5	e_i

ارسم الرواسب e_i مقابل X_i ما هي الاستنتاجات التي يمكنك استخلاصها من الرسم؟

(٤-١٢) تصرح طالبة أنها لاتستيطع فهم لماذا يسمى مجموع المربعات المعرف في (4.11) بمجموع مربعات خطأ بحت "إذ تبدو الصيغة كواحدة من صيغ مجاميع المربعات الاعتيادية". وضّح.

(٤-١٣) بالعودة إلى **صيانة الحاسبات** مسألة (٢-١٨) إليك بعض النتائج الحسابية الإضافية:

$$SSE = 321.4 \text{ و } SSR = 16,182.6$$

ا ـ في اختبار F لنقص توفيق دالة انحدار خطّية، ما هي النتائج في الفرضية البديلة؟

ب ـ قم بالاختبار الموضح في الجزء (أ) اضبط مخاطرة التورط في خطأ من النوع الأول عند 0.05. أذكر قاعدة القرار والنتيجة.

جـ ـ هل يكشف اختبارك في الجزء (ب) عن انحرافات أخرى عن نموذج الانحدار (3.1)، مثل نقص في ثبات التباين أو نقص في طبيعية حدود الخطأ؟ هل يمكن أن تتأثر نتائج الاختبار بمثل هذه الانحرافات؟ ناقش.

(٤-١٤) بالعودة إلى مسألة **صلابة البلاستيك** (٢-٢٠).

ا ـ قم باختيار F لتحديد ما إذا كان هناك نقص توفيق لدالة انحدار خطية أم لا. استخدم مستوى معنوية 0.01. اذكر البديلين وقاعدة القرار والنتيجة.

ب ـ هل هناك أية فائدة تُرجى لتساوي التكرارات عند كل من مستويات X؟ هل هناك أية مساوىء ؟

جـ ـ عندما يؤدي الاختبار في (أ) إلى نتيجة أن دالة الانحدار غير خطّية فهل يشير ذلك الاختبار إلى دالة الانحدار المناسبة؟ كيف نمضي من هنا ؟

(٤ـ١٥) **تركيز محلول**. درس كيميائي تركيز محلول Y فوق فترة زمنية X. وقد قام بإعداد خمسة عشر محلولا متماثلا. ثم قسّم الخمسة عشر محلولا بصورة عشوائية إلى خمس مجموعات من ثلاثة وقيست المجموعات الخمس، على الترتيب بعد 1، 3، 5، 7 و9 ساعات. وكانت النتائج كما يلي:

i	1	2	3	4	5	6	7
X_i	9	9	9	7	7	7	5
Y_i	.07	.09	.08	.16	.17	.21	.49

i	8	9	10	11	12	13	14	15
X_i	5	5	3	3	3	1	1	1
Y_i	.58	.53	1.22	1.15	1.07	2.84	2.47	3.10

ا ـ قم بتوفيق دالة انحدار خطّية.

ب ـ قم باختبار F لتحديد ما إذا كان هناك نقص توفيق لدالة الانحدار الخطّية أم لا؛ مستخدما $\alpha = 0.025$. اذكر البديلين وقاعدة القرار والنتيجة.

جـ ـ عندما يقود الاختبار في الجزء (ب) إلى وجود نقص توفيق في دالة الانحدار الخطية فهل يشير ذلك الاختبار إلى دالة الانحدار المناسبة ؟ اشرح.

(٤ـ١٦) بالعودة إلى مسألة **تركيز محلول** (٤–١٥).

ا ـ قم بإعداد رسم انتشار للبيانات. ماهو التحويل الذي يمكن تجريبه على Y بالاستناد إلى الأنماط المعروضة في الشكل (٤–١٥) وذلك للحصول على تباين ثابت وعلى الخطية؟.

ب ـ استخدم طريقة بوكس ـ كوكس والمعايرة (4.29) لإيجاد تحويل قوى مناسب، احسب SSE من أجل $\lambda = -0.2, -0.1, 0, 0.1, 0.2$. ما هو تحويل Y المقترح؟

جـ ـ استخدم التحويل $Y' = log_{10}Y$ وأوجد دالة الانحدار المقدّرة للبيانات بعد التحويل.

د ـ ارسم خط الانحدار المقدّر والبيانات بعد التحويل. هل يبدو خط الانحدار توفيقا جيدا للبيانات بعد التحويل ؟.

هـ - أوجد الرواسب وارسمها مقابل القيم التوفيقية. قم أيضا بإعداد رسم احتمال طبيعي. ماذا توضّح رسومك ؟.

و - عبّر عن دالة الانحدار المقدّرة بالوحدات الأصلية.

(٤-١٧) **نمو المبيعات.** درس باحث تسويق المبيعات السنوية لمُنتج طُرح في السوق منذ 10 سنوات. وكانت البيانات كالتالي، حيث X السنة (مرمّزة) وY المبيعات بآلاف الوحدات:

i	1	2	3	4	5	6	7	8	9	10
X_i	0	1	2	3	4	5	6	7	8	9
Y_i	98	135	162	178	221	232	283	300	374	395

ا - قم بإعداد رسم انتشار للبيانات، هل تبدو العلاقة الخطّية مناسبة هنا ؟

ب - استخدم طريقة بوكس - كوكس والمعايرة (4.29) لايجاد تحويل قوة مناسب لـ λ. احسب SSE من أجل $\lambda = 0.3, 0.4, 0.5, 0.6, 0.7$ ما هو تحويل Y المقترح ؟

جـ - استخدم التحويل $Y' = \sqrt{Y}$ وأوجد دالة الانحدار الخطية المقدَّرة للبيانات بعد التحويل.

د - ارسم خط الانحدار المقدَّر والبيانات بعد التحويل - هل يبدو خط الانحدار توفيقا جيدا للبيانات بعد التحويل؟.

هـ - أوجد الرواسب وارسمها مقابل القيم التوفيقية قم أيضا بإعداد رسم احتمال طبيعي. ماذا توضح رسومك؟.

و - عبّر عن دالة الانحدار المقدَّرة بالوحدات الأصلية.

(٤-١٨) **أخطاء الاستجابة.** في دراسة على نطاق ضيق لأخطاء الاستجابة عند تذكر النفقات الحاصلة خلال أحدث رحلة، تم الحصول على البيانات التالية عن أخطاء الاستجابة Y وعدد الأشهر منذ آخر رحلة صيد (X):

i	1	2	3	4	5	6
X_i	12	5	1	8	1	3
Y_i	-94	-68	-41	-78	-36	-55

i	7	8	9	10	11	12
X_i	15	9	7	2	10	4
Y_i	0106	-90	-78	-49	-85	-76

ا - قم بإعداد رسم انتشار للبيانات. هل تبدو العلاقة الخطّية مناسبة هنا؟
أيها أفضل هنا تحويل X أو تحويل Y؟ لماذا؟

ب - استخدم التحويل $X' = \sqrt{X}$ وأوجد دالة الانحدار الخطّية المقدَّرة للبيانات بعد التحويل.

جـ - ارسم خط الانحدار المقدَّر والبيانات بعد التحويل هل يبدو خط الانحدار توفيقا جيدا للبيانات بعد التحويل؟

د - أوجد الرواسب وارسمها مقابل القيم التوفيقية. وقم أيضا بإعداد رسم احتمال طبيعي. ماذا توضح رسومك ؟.

هـ - عبّر عن دالة الانحدار المقدَّرة بالوحدات الأصلية.

تمارين

(٤-١٩) قام طالب بتوفيق دالة انحدار خطّية كواجب دراسي. وفيما يلي بعض النتائج:

5	4	3	2	1	i
53	28	42	17	35	X_i
40	32	32	29	42	$\hat{Y}_i$
13	-4	10	-12	-7	e_i

رسم الطالب الرواسب e_i مقابل Y_i ووجد علاقة موجبة. وعندما رسم الرواسب مقابل القيم التوفيقية $\hat{Y}_i$ لم يجد أي علاقة. لماذا توجد هذه المفارقة، وأي الرسمين له معنى أكثر ؟

(٤-٢٠) إذا كانت حدود الخطأ في نموذج انحدار $N(0,\sigma^2)$، ومستقلة، فماذا يمكن القول عن حدود الخطأ بعد استخدام التحويل $X' = 1/X$ ؟ هل تبقى الحالة نفسها بعد استخدام التحويل $Y' = 1/Y$ ؟.

(٤-٢١) استنبط النتيجة في (4.24).

(٤-٢٢) مستخدما النظريات (1.27)، (1.38) و(1.39) أثبت أن $E\{MSPE\}=\sigma^2$ لنموذج انحدار الخطأ الطبيعي.

(٤-٢٣) النموذج قيد الدراسة هو نموذج انحدار خطّي بمقطوع $\beta_0 = 0$ وقد تم الحصول على بيانات تتضمن تكرارات. اعرض النموذجين التام والمخفض

لاختبار صلاحية دالة الانحدار قيد الدراسة. ما هي درجات الحرية المتعلقة بالنموذجين التام والمخفض إذا كان $n = 20$ و $c = 10$ ؟

مشاريع

(٤-٢٤) **ضغط الدم.** تم الحصول على البيانات التالية في دراسة للعلاقة بين ضغط الدم الإنبساطي Y والعمر X لصبية تتراوح أعمارهم من 5 إلى 13 سنة.

8	7	6	5	4	3	2	1	i
6	12	12	13	7	11	8	5	X_i
60	90	69	75	64	74	67	63	Y_i

ا ـ مفترضا أن نموذج انحدار الخطأ الطبيعي (3.1) مناسب، أوجد دالة الانحدار المقدَّرة وارسم الرواسب e_i مقابل X_i . ماذا يبين رسمك للرواسب؟

ب ـ احذف المشاهدة 7 من البيانات وأوجد دالة الانحدار المقدَّرة بناء على المشاهدات السبع المتبقية. قارن دالة الانحدار المقدَّرة هذه بتلك التي حصلت عليها في الجزء (أ). ماذا يمكنك أن تستنتج عن تأثير المشاهدة 7؟.

جـ ـ مستخدما دالة الانحدار التوفيقية في الجزء (ب)، أوجد 99 بالمائة فترة تنبؤ لمشاهدة جديدة Y عند $X = 12$. هل تقع المشاهدة Y_7 خارج فترة التنبؤ هذه ؟ مادلالة ذلك؟

(٤-٢٥) عُد إلى مجموعة بيانات **SMSA** والمشروع (٢-٤٢). لكل من نماذج الانحدار التوفيقية الثلاثة، أوجد الرواسب وقم بإعداد رسم رواسب مقابل X ورسم احتمال طبيعي، لخّص استنتاجاتك. هل نموذج الانحدار الخطي (3.1) أكثر ملاءمة لإحدى الحالات من الأخريات؟

(٤-٢٦) بالعودة إلى مجموعة بيانات **SMSA** ومشروع (٢-٤٣). لكل منطقة جغرافية، أوجد الرواسب، وقم بإعداد رسم رواسب مقابل X ورسم احتمال طبيعي. هل يبدو أن للمناطق الأربع تباينات خطأ متشابهة ؟ ماهي الاستنتاجات الأخرى التي يمكنك استخلاصها من الرسوم؟

(٤-٢٧) بالعودة إلى مجموعة بيانات **SENIC** ومشروع (٢-٤٤).

ا ـ لكل من نماذج الانحدار التوفيقيـة الثلاثـة، أوجـد الرواسـب وقـم بـإعداد رسم رواسب مقابل X ورسم احتمال طبيعي، لخّص استنتاجاتك. هل يبدو نموذج الانحدار الخطّي (3.1) أكثر ملاءمـة لإحـدى الحـالات منـه للحالات الأخرى ؟

ب ـ أوجد دالة الانحدار التوفيقية للعلاقة بين مدة الإقامـة ومخـاطر العـدوى بعد حذف المشاهدتين:

$47(X_{47} = 6.5,\ Y_{47} = 19.56)$ و $112(X_{112} = 5.9,\ Y_{112} = 17.94)$.

من دالة الانحدار التوفيقية هذه. أوجد 95 بالمائة فترتي تنبؤ منفصلتين لمشاهدتي Y جديدتين عند $X = 6.5$ و $X = 5.9$ على الترتيب. هـل تقـع المشاهدتان Y_{47} و Y_{112} خارج فترتي التنبؤ هاتين ؟ ناقش دلالة ذلك.

(٤-٢٨) بـالعودة إلى مجموعـة بيانـات **SENIC** ومشـروع (٢-٤٥). ولكـل منطقـة جغرافية، أوجد الرواسب وقم بإعداد رسم رواسب مقابل X ورسم احتمـال طبيعـي. هـل يبـدو للمنـاطق الأربـع تباينـات خطـأ متشـابهة ؟ مـاهي الاستنتاجات الأخرى التي تستخلصها من رسومك؟.



استقراءات متزامنة ومواضيع أخرى في تحليل الانحدار

في هذا الفصل نتابع مواضيع متنوعة في تحليل الانحدار الخطّي البسيط. ويتعلق العديد من المواضيع بمسألة كيفية القيام باستقراءات متزامنة من مجموعة مشاهدات العينة نفسها.

(٥-١) التقدير المشترك لـ β_0 و β_1

الحاجة لتقدير مشترك

يقوم محلل أبحاث تسويق بدراسة العلاقة بين مستوى الإعلان (X) والمبيعات Y وقد كانت هناك بعض المشاهدات بدون إعلان ($X = 0$) بينما اختلف مستوى الإعلان في المشاهدات الأخرى. واقترح رسم الانتشار علاقة خطية ضمن مدى مستويات نفقات الإعلان المدروسة. ويرغب المحلل استخلاص استقراءات حول كل من الجزء المقطوع β_0 والميل β_1. ويمكن للمحلل استخدام طرق الفصل الثالث لإقامة 95 بالمائة فترتي ثقة منفصلتين لكل من β_0 و β_1. والمشكلة هنا هي أن هاتين الفترتين لاتقدّمان 95 بالمائة معامل ثقة للقراران لاستنتاج أن فترتي الثقة حول β_0 و β_1 صحيحتان معا أو في الوقت نفسه. ولو كان الاستقراءان مستقلين فإن احتمال صحتهما معا هو $(0.95)^2$ أو 0.9025 فقط. وعلى أي حال، فإن الاستقراءين ليسا مستقلين إذ يأتيان من مجموعة بيانات العينة نفسها، مما يجعل تحديد احتمال كونهما صحيحين معا أكثر صعوبة بكثير.

وكثيرا ما يتطلب تحليل البيانات، سلسلة من التقديرات (أو الاختبارات) ويريد المحلل أن يكون مطمئنا إلى صحة مجموعة التقديرات (أو الاختبارات) كافة. وسندعو مجموعة التقديرات (أو الاختبارات) قيد الاعتبار *عائلة التقديرات* (أو الاختبارات). وتتألف

العائلة في توضيحنا من تقديرين، لـ β_0 و β_1، على الترتيب. ومن ثم نميز بين معامل ثقة عبارة ومعامل ثقة عائلة. ***فمعامل ثقة عبارة*** هو النوع المعتاد لمعامل الثقة الذي نوقش سابقا، والذي يشير إلى نسبة التقديرات الصحيحة التي نحصل عليها عندما نكرر أخذ عينات، ونحسب فترة الثقة المحددة لكل عينة. ***ومعامل ثقة عائلة***، من جهة أخرى، يشير إلى نسبة العائلات من التقديرات التي تكون جميع تقديراتها الفردية صحيحة وذلك عندما نكرر أخذ عينات ونحسب من كل عينة فترات الثقة لكل فرد من أفراد العائلة. وهكذا فإنه قبل القيام بالمعاينة، يقابل معامل ثقة عائلة احتمال أن جميع عبارات العائلة ستكون صحيحة في آن واحد.

ولمزيد من التوضيح لمعنى معامل ثقة عائلة، لنعد إلى التقدير المشترك لـ β_0 و β_1 فيشير معامل ثقة عائلة، ولنقل 0.95 إلى أنه إذا تكرر أخذ عينات وحسبنا تقديري فترة لـ β_0 و β_1 من كل عينة بالطرق المحددة ذاتها، فإن 95 بالمائة من العينات ستؤدي إلى عائلة تقديرات تكون فيها فترتا الثقة صحيحتين. ومن أجل 5 بالمائة من العينات ستكون إما واحدة من الفترتين أو كلتاهما غير صحيحة.

فترات الثقة المشتركة لبونفيرّوني

من الواضح أن الطريقة التي تقدم معامل ثقة عائلة تكون في العادة مرغوبة جدا، لأنها تسمح للمحلل بوضع النتائج المنفصلة مع بعضها في مجموعة متكاملة من القرارات، مع الاطمئنان إلى أن مجموعة التقديرات بكاملها صحيحة. وطريقة بونفرّوني لتطوير فترات ثقة مشتركة مع معامل ثقة عائلة محدد هي طريقة سهلة جدا: يعدَّل معامل ثقة كل عبارة ليكون أعلى من $1 - \alpha$ وبحيث يكون معامل ثقة العائلة، على الأقل $1 - \alpha$، والطريقة هي طريقةعامة يمكن تطبيقها في العديد من الحالات، كما سنرى، وليس فقط على التقدير المشترك لـ β_0 و β_1. ونوضح هنا طريقة بونفرّوني مطبقة على تقدير β_0 و β_1 بصورة مشتركة.

نبدأ بحدي الثقة الاعتياديين لـ β_0 و β_1 بمعاملي ثقة عبارة يساوي $1-\alpha$ لكل منهما. وهذه الحدود هي:

$$b_0 \pm t(1-\alpha/2; n-2)s\{b_0\}$$
$$b_1 \pm t(1-\alpha/2; n-2)s\{b_1\}$$

ومن ثَمَّ نتساءل: ماهو احتمال أن تكون كلا المجموعتين من الحدود صحيحتين؟. لنرمز بـ A_1 لحادثة أن فترة الثقة الأولى لا تغطي β_0 ولنرمز بـ A_2 لحادثة أن فترة الثقة الثانية لا تغطي β_1 فنعلم أن:

$$P(A_1) = \alpha \qquad P(A_2) = \alpha$$

وتعرض نظرية الاحتمال (1.6):

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

وبالتالي:

$$1 - P(A_1 \cup A_2) = 1 - P(A_1) - P(A_2) + P(A_1 \cap A_2) \qquad (5.1)$$

والآن من نظريتي الاحتمال (1.9) و(1.10)، نجد:

$$1 - P(A_1 \cup A_2) = P(\overline{A_1 \cup A_2}) = P(\overline{A}_1 \cap \overline{A}_2))$$

$P(\overline{A}_1 \cap \overline{A}_2) \geq 0$ هو احتمال أن فترتي الثقة كليهما صحيحتان. وهكذا نجد من (5.1) أن:

$$P(\overline{A}_1 \cap \overline{A}_2) = 1 - P(A_1) - P(A_2) + P(A_1 \cap A_2) \qquad (5.2)$$

وحيث إن $P(A_1 \cap A_2) \geq 0$، فنحصل من (5.2) على متباينة بونفرّوني:

$$P(\overline{A_1} \cap \overline{A_2}) \geq 1 - P(A_1) - P(A_2) \qquad (5.3)$$

والتي تصبح في حالتنا:

$$P(\overline{A_1} \cap \overline{A_2}) \geq 1 - \alpha - \alpha = 1 - 2\alpha \qquad (5.3a)$$

وهكذا، فإنه إذا قُدِّرت β_0 و β_1 بصورة منفصلة، ولنقل، بـ 95 بالمائة فترة ثقة فإن متباينة بونفرّوني تضمن لنا بمعامل ثقة عائلي يساوي 90 بالمائة على الأقل أن الفترتين الناتجتين من العينة نفسها صحيحتان معا.

ويمكننا بسهولة استخدام متباينة بونفرّوني (5.3) للحصول على معامل ثقة عائلي يساوي، على الأقل $\alpha - 1$ لتقدير β_0 و β_1. ونقوم بذلك من خلال تقدير β_0 و β_1 بصورة منفصلة، وبمعامل ثقة عبارة يساوي $1-\alpha/2$ لكل منهما، وهذا يعطي حد

بونفرّوني $1-\alpha = 1-\alpha/2-\alpha/2$. وهكذا يكون الـ $1-\alpha$ حدي ثقة عائليين لـ β_0 و β_1، وتدعى غالبا *مجموعة ثقة*، مساوين وفقا لطريقة بونفيروني

$$\begin{aligned} b_0 \pm Bs\{b_0\} \\ b_1 \pm Bs\{b_1\} \end{aligned} \tag{5.4}$$

حيث:

$$B = t(1-\alpha/4; n-2) \tag{5.4a}$$

لاحظ أن معامل ثقة عبارة مساوٍ لـ $1-\alpha/2$ يتطلب المئين $100(1-\alpha/4)$ لتوزيع t وذلك لفترة ثقة ثنائية الجانب.

مثال

في مثال حجم الدفعة لشركة وستوود، يتطلب 90 بالمائة فترة ثقة عائلة لـ β_0 ، β_1 و $B = t(1-.10/4;8) = t(.975;8) = 2.306$ ولدينا من الفصل الثالث:

$$\begin{aligned} b_0 &= 10.0 \qquad & s\{b_0\} &= 2.50294 \\ b_1 &= 2.0 \qquad & s\{b_1\} &= .04697 \end{aligned}$$

وبالتالي يكون حدا الثقة لـ β_0 و β_1 على الترتيب $10.0 \pm 2.306(2.50294)$ و$2.0 \pm 2.306\ (0.04697)$ ، وفترتا الثقة المشتركة هما:

$$\begin{aligned} 4.2282 \le \beta_0 \le 15.7718 \\ 1.8917 \le \beta_1 \le 2.1083 \end{aligned}$$

وهكذا، نستنتج أن تقع β_0 بين 4.23 و15.77 وتقع β_1 بين 1.89 و 2.11. وبمعامل ثقة عائلي 0.90، على الأقل، نستنتج أن الطريقة تقود إلى تقديري فترة صحيحين.

تعليقات

١- نكرر هنا أن الـ $1-\alpha$ معامل ثقة عائلي لبونفرّوني هو، في الواقع، حد أدنى لمعامل الثقة العائلي الفعلي (وهو عادة غير معروف). وإلى الحد الذي يميل فيه تقديرا الفترة غير الصحيحين لـ β_0 و β_1 إلى الظهور معا في عائلة التقديرات، فإن العائلة من العبارات تنجو إلى أن تكون صحيحة بأكثر من $100(1-\alpha)\%$ من المرّات. وبسبب هذه الطبيعة المحافظة لطريقة بونفرّوني، فإنه غالبا ماتتحدد معاملات الثقة العائلية عند مستويات أدنى (مثلا 90 بالمائة) مما لو كان المطلوب معامل ثقة لتقدير بمفرده.

٢- يمكن تمديد متباينة بونفرّوني ($5.3a$) بسهولة إلى g من فترات الثقة المتزامنة بمعامل ثقة عائلي $1 - \alpha$.

$$P\left(\bigcap_{i=1}^{g} \overline{A}_i\right) \geq 1 - g\alpha \qquad (5.5)$$

وهكذا إذا أردنا g من التقديرات بفترة، بمعامل ثقة عائلي α - 1 يكفي إقامة كل تقدير بفترة بمعامل ثقة عبارة α/g - 1.

٣- من أجل معامل ثقة عائلي معطى، فإنه كلما كبر عدد فترات الثقة في العائلة أصبح العامل B أكبر، مما قد يجعل بعض أو كل فترات الثقة من الاتّساع بحيث لافائدة ترجى منها. وفائدة طريقة بونفرّوني هي في العادة أكبر مايمكن عندما لايكون عدد التقديرات المتزامنة كبيرا جدا.

٤- ليس من الضروري في طريقة بونفرّوني أن يكون لفترات الثقة معاملات ثقة العبارة نفسها. إذ يمكن استخدام معاملات ثقة عبارة تختلف باختلاف أهمية كل تقدير. فمثلا، في توضيحنا السابق، قد تُقدَّر β_0 بـ 92 بالمائة فترة ثقة و β_1 بـ 98 بالمائة فترة ثقة. ومن (5.3) نجد أن معامل ثقة العائلة سيبقى، على الأقل، 90 بالمائة.

٥- يمكن استخدام فترات الثقة المشتركة بصورة مباشرة للقيام باختبار. ولتوضيح هذا الاستخدام، افترض أن مهندسا صناعيا يعمل لدى شركة وستوود قد وضع نظرية تزعم أن لدالة الانحدار مقطوعا يساوي 13.0 وميلا يساوي 2.5. وفي حين تقع 13.0 ضمن فترة ثقة β_0 فإن 2.5 لاتقع في فترة ثقة β_1. وهكذا لاتكون توقعات المهندس النظرية صحيحة عند مستوى أهمية عائلي $\alpha = 0.10$.

٦- المقدّران β_0 و β_1 مرتبطان عادة، ولكن حدي ثقة بونفرّوني المتزامنين في (5.4) يميزان هذا الارتباط فقط من خلال الحد الأعلى لمعامل ثقة العائلة. ويمكن إثبات أن التغاير بين b_0 وb_1 هو:

$$\sigma\{b_0, b_1\} = -\overline{X}\sigma^2\{b_1\} \qquad (5.6)$$

ونلاحظ أنه إذا كان $\overline{X}$ موجبا فإن b_0 و b_1 يرتبطان سلبيا مما يؤدي إلى أنه إذا كان المقدر b_1 كبيرا جدا فإن المقدَّر b_0 يكون في الغالب صغيرا جدا والعكس بالعكس.

وفي مثال شركة وستوود لدينا $\bar{X}=50$ ، وبالتالي فإن التغاير سالب. وهذا يتضمن أن المقدرين b_0 و b_1 يميلان هنا إلى الجنوح في اتجاهين متضادين. ونتوقع هذا بالبداهة. وبما أن النقاط الملحوظة $(X_i\ ,\ Y_i)$ تقع في الربع الأول (انظر الشكل (٢-٢) أ) فمن المنتظر أنه إذا كان خط الانحدار التوفيقي شديد الانحدار(b_1 تبالغ في تقدير β_1) فمن المحتمل جدا أن يكون المقطوع منخفضا جدا (b_0 تقدير بالنقصان لـβ_0)، والعكس بالعكس.

وعندما يكون المتغير المستقل $X_i-\bar{X}$ ، كما في النموذج البديل (2.6) فلا يكون b_0^* و b_1 وفقا لـ (5.6) مرتبطين لأن متوسط المشاهدات $X_i-\bar{X}$ صفر.

(٥ ـ ٢) تقدير متزامن لمتوسط الاستجابات

نرغب عادة في تقدير متوسطات الاستجابات عند عدد من مستويات X وذلك من بيانات العينة نفسها فقد ترغب شركة وستوود، على سبيل المثال، تقدير متوسط عدد ساعات العمل لدفعات من 30، 55 و 80 قطعة. ونعلم فيما سبق كيف نقوم بذلك لأي مستوى من مستويات X وذلك بمعامل ثقة عبارة معطى، وسنناقش الآن طريقتين لتقدير متزامن لمتوسطات الاستجابة بمعامل ثقة عائلة، وبحيث يتوافر ضمان معروف لكون جميع تقديرات متوسطات الاستجابة صحيحة. والطريقتان هما طريقة وركنج - هوتّلنج وطريقة بونفرّوني.

وسبب الاهتمام بمعامل ثقة عائلة هو أنه ليس ضروريا أن التقديرات المنفصلة لـ$E\{Y_h\}$ عند مستويات X_h مختلفة، جميعها صحيحة أو جميعها غير صحيحة، وذلك بالرغم من أنها جميعها مرتكزة على بيانات العينة نفسها وخط الانحدار التوفيقي نفسه. وقد يكون مركّب خطأي المعاينة لـ b_0 و b_1 بحيث يجعل تقديرات الفترة لـ $E\{Y_h\}$ صحيحة فوق مدى من مستويات X وغير صحيحة فيما عدا ذلك.

طريقة وركنج ـ هوتلّنج

تنطبق طريقة وركنج ـ هوتلّنج عندما تتألف عائلة التقديرات من جميع مستويات X الممكنة. وفي العادة نرغب، بالطبع، تقدير متوسط الاستجابة لعدد محدود فقط من مستويات X. وكنتيجة لذلك، فإن معامل الثقة العائلي الفعلي عند تطبيق

طريقة ووركنج ـ هوتلّنج لتقدير عدد محدّد فقط من مستويات الاستجابة سيكون، في واقع الأمر، أكبر من معامل الثقة المزعوم.

حدود الثقة المتزامنة لِ g متوسط استجابة $E\{Y_h\}$ بطريقة ووركنج هوتلّنج هي من الشكل:

$$\hat{Y}_h \pm Ws\{\hat{Y}\} \tag{5.7}$$

حيث:

$$W^2 = 2F(1 - \alpha\,; 2, n - 2) \tag{5.7a}$$

و $\hat{Y}_h$ و $s\{\hat{Y}_h\}$ معرفان في (3.27) و(3.29)، على الترتيب.

مثال. في مثال حجم دفعة شركة وستوود، نريد عائلة تقديرات لمتوسطات عدد ساعات العمل عند المستويات التالية لحجم الدفعة: 30، 55 و80. ومعامل ثقة العائلة المطلوب هو 0.90. وقد حصلنا في الفصل الثالث، على $\hat{Y}_h$ و $s\{\hat{Y}_h\}$ لـ $X_h = 55$. وبطريقة مشابهة نحصل على النتائج المطلوبة للدفعتين الأخيرتين. ونلخّصها هنا بدون توضيح الحسابات.

X_h	$\hat{Y}_h$	$s\{\hat{Y}_h\}$
30	70.0	1.27764
55	120.0	0.89730
80	170.0	1.65387

ولمعامل ثقة عائلة يساوي 0.90، نحتاج إلى $F(0.90; 2,8) = 3.11$ وبالتالي:

$$W^2 = 2(3.11) = 6.22 \qquad W = 2.494$$

ونستطيع الآن الحصول على فترات ثقة لمتوسطات عدد ساعات – العمل عند $X_h = 30$، $X_h = 55$ و $X_h = 80$:

$$66.8 = 70.0 - 2.494(1.27764) \le E\{Y_h\} \le 70.0 + 2.494(1.27764) = 73.2$$
$$117.8 = 120.0 - 2.494(0.89730) \le E\{Y_h\} \le 120.0 + 2.494(0.89730) = 122.2$$
$$165.9 = 170.0 - 2.494(1.65387) \le E\{Y_h\} \le 170.0 + 2.494(1.65387) = 174.1$$

وبمعامل ثقة عائلة 0.90 نستنتج أن متوسط عدد ساعات ـ العمل لدفعات من 30 قطعة يقع بين 66.8 و 73.2. ولدفعات من 55 قطعة يقع بين 117.8 و122.2 أما لدفعات من 80 قطعة فيقع بين 165.9 و174.1. ويقدم معامل ثقة العائلة 0.90 ضمانا

بأن الطريقة تقود إلى أن التقديرات جميعها، في هذه العائلة من التقديرات، صحيحة.

طريقة بونفرّوني

طريقة بونفرّوني التي ناقشناها سابقا حول التقدير المتزامن لـِ β_0 و β_1، هي طريقة عامة تماما. ولبناء عائلة من فترات الثقة لمتوسطات الاستجابة عند مستويات مختلفة لـX، نحسب حدي الثقة المعتادين لمتوسط استجابة بمفرده $E\{Y_h\}$ كما هو معطى في (3.32)، ونعدل معامل ثقة العبارة لينتُج معامل ثقة العائلة المحدّد.

وعندما يُراد تقدير $E\{Y_h\}$ لِـ g من المستويات X_h، بمعامل ثقة عائلي $1 - \alpha$ فإن حدي ثقة بونفيروني هما:

$$\hat{Y}_h \pm Bs\{\hat{Y}_h\} \qquad (5.8)$$

حيث:

$$B = t(1 - \alpha/2g; n-2) \qquad (5.8a)$$

g عدد فترات الثقة في العائلة

مثال. تتطلب تقديرات متوسطات عدد ساعات – العمل بطريقة بونفرّوني لدفعات حجومها 30 ، 55 و80 قطعة وبمعامل ثقة عائلة 0.90، البيانات نفسها التي تتطلبها في طريقة ووركنج – هوتلّنج المعروضة أعلاه. وبالإضافة إلى ذلك نحتاج إلى:

$$B = t[1 - 0.10/2(3); 8] = t[0.983; 8]$$

وباستيفاء خطي في الجدول (أ-٢) نجد $t(0.983; 8) = 2.56$ (انظر التعليق ٤ التالي).

وهكذا نحصل على فترات الثقة لمتوسطات عدد ساعات العمل لحجوم الدفعات $X_h = 30$، $X_h = 55$ و $X_h = 80$ بـ 90 بالمائة معامل ثقة عائلي، كالتالي:

$$66.7 = 70.0 - 2.56(1.27764) \leq E\{Y_h\} \leq 70.0 + 2.56(1.27764) = 73.3$$
$$117.7 = 120.0 - 2.56(.89730) \leq E\{Y_h\} \leq 120.0 + 2.56(.89730) = 122.3$$
$$165.8 = 170.0 - 2.56(1.65387) \leq E\{Y_h\} \leq 170.0 + 2.56(1.65387) = 174.2$$

تعليقات

١- في هذه الحالة يكون حدا ثقة ووركنج – هوتلّنج أضيق قليلا من حدي بونفرّوني. وفي حالات أخرى حيث يكون عدد العبارات صغيرا، قد يكون حدا بونفرّوني أضيق. وللعائلات الكبيرة يكون حدا ثقة ووركنج ـ هوتلّنج أضيق دائما،

لأن W في ($5.7a$) تبقى نفسها لأي عدد من العبارات في العائلة، بينما تصبح B في (5.8a) كبيرة كلما ازداد عدد العبارات. وعند التطبيق، حالما نحدد معامل ثقة العائلة، يمكن حساب المضاعفات W وB وتحديد الطريقة التي تقود إلى حدي ثقة أضيق.

٢- تقدم كل من طريقتي ووركينج ـ هوتلّنج وبونفرّوني للتقديرات المتعددة لمتوسطات الاستجابة حدود دنيا لمعامل ثقة العائلة الفعلي.

٣- في بعض الأحيان لاتُعرف مقدما مستويات المتغير المستقل التي نقدّر عندها متوسطات الاستجابة. بل يجري تحديدها مع سير التحليل. وفي مثل هذه الحالات يُفضل استخدام أسلوب ووركينج ـ هوتلّنج لأن العائلة لهذا الأسلوب تشمل كل مستويات X الممكنة.

٤- لإيجاد مئينات غير مجدولة لتوزيع t، يُعطي الاستيفاء الخطّي في جدول (أ-٢)، عادة، تقريبات جيدة بصورة معقولة طالما كان عدد درجات الحرية غير صغير. وفي توضيحنا لطريقة بونفرّوني احتجنا $t(0.983;8)$ ونعلم من جدول (أ-٢) أن:

$$t(.985;8) = 2.634 \qquad t(.980;8) = 2.449$$

ويعطي الاستيفاء الخطّي:

$$t(0.983;8) = 2.449 + \left(\frac{0.983 - 0.980}{0.985 - 0.980}\right)(2.634 - 2.449) = 2.56$$

وتقدم كثير من حزم الحاسب الآلي مئينات التوزيع t وتوزيعات أخرى، مما يغني عن الاستيفاء من الجداول.

(٥ ـ ٣) فترات تنبؤ متزامنة لمشاهدات جديدة

نعتبر الآن التنبؤ المتزامن لـ g مشاهدة Y جديدة في g محاولة مستقلة عند g مستوى مختلف لـ X. ولتوضيح هذا النوع من التطبيق، دعنا نفترض أن شركة وستوود تخطط لإنتاج الدفعات الثلاث القادمة بحجوم 30، 55 و80 قطعة، وترغب التنبؤ بساعات العمل لكل من هذه الدفعات بمعامل ثقة عائلي 0.95.

وسنستعرض هنا طريقتين، طريقة شفّيه وطريقة بونفرّوني. وكلتاهما تستفيد من نوع الحدود نفسه، كما هو في حالة التنبؤ بمشاهدة واحدة، والمعطاة في (5.35) ويتغير

فقط مضاعف الانحراف المعياري المقدّر. وتستخدم طريقة شفّيه توزيع F، في حين تستخدم طريقة بونفرّوني توزيع t وحدود التنبؤ المتزامنة لـ g تنبؤ بطريقة شفّيه بمعامل ثقة عائلي $1 - \alpha$ هي:

$$\hat{Y}_h \pm Ss\{Y_{h(new)}\} \qquad (5.9)$$

حيث:

$$S^2 = gF(1 - \alpha; g, n - 2) \qquad (5.9a)$$

وبطريقة بونفرّوني، تكون الـ $1 - \alpha$ حدود تنبؤ متزامنة:

$$\hat{Y}_h \pm Bs\{Y_{h(new)}\} \qquad (5.10)$$

حيث:

$$B = t(1 - \alpha/2g; n - 2) \qquad (5.10a)$$

ونستطيع حساب المضاعفين S و B لرؤية أي الطريقتين تقدم حدي تنبؤ أضيق.

ولمثالنا، لدينا:

$$S^2 = 3F(.95; 3, 8) = 3(4.07) = 12.21 \qquad S = 3.49$$
$$B = t[1 - 0.05/2(3); 8] = t(0.992; 8) = 3.04$$

ولذلك سنستخدم هنا طريقة بونفرّوني. ومن نتائج سابقة، نجد (الحسابات غير موضحة):

X_h	$\hat{Y}_h$	$s\{Y_{h(new)}\}$	$Bs\{Y_{h(new)}\}$
30	70.0	3.02198	9.18682
55	120.0	2.88187	8.76088
80	170.0	3.19926	9.72575

وحدود التنبؤ المتزامنة للدفعات الثلاث القادمة، بمعامل ثقة عائلي 0.95 حيث $X_h = 30$، $X_h = 55$ و $X_h = 80$ هي:

$$60.8 = 70.0 - 9.18682 \le Y_{h(new)} \le 70.0 + 9.18682 = 79.2$$
$$111.2 = 120.0 - 8.76088 \le Y_{h(new)} \le 120.0 + 8.76088 = 128.8$$
$$160.3 = 170.0 - 9.72575 \le Y_{h(new)} \le 170.0 + 9.72575 = 179.7$$

وبمعامل ثقة عائلة لايقل عن 0.95، يمكننا التنبؤ بأن ساعات العمل لأشواط الإنتاج الثلاثة القادمة ستكون ضمن الحدين أعلاه.

تعليقات

١- فترات التنبؤ المتزامنة لـ g مشاهدة Y جديدة عند g مستوى مختلف لـ X بـ $1-\alpha$ معامل ثقة عائلي أوسع من فترات الثقة المفردة المقابلة (3.35). وعندما لايكون

عدد التنبؤات كبيرا، على أي حال، فإن الفرق في عرض الفترتين يبقى معتدلا. وعلى سبيل المثال، كانت 95 بالمائة فترة تنبؤ مفردة في مثال شركة وستوود ستستخدم المضاعف t حيث $t(0.975;8) = 2.306$ وهو أصغر بقليل من المضاعف $B=3.04$ للتنبؤات الثلاثة معا.

٢- لاحظ أن كلا من B و S تصبح أكبر مع ازدياد g. خلافا لما نجده في التقديرات المتزامنة لمتوسطات الاستجابة، حيث تصبح B الأكبر إلا أن W لا تزداد. عندما تكون g كبيرة، قد يصبح المضاعفان B و S كبيرين إلى الحد الذي يجعل فترات التنبؤ من الاتساع بحيث تصبح غير ذات جدوى. وعندئذ يمكن الاستفادة من تقانات تقدير متزامنة أخرى، مما نجده في المرجع [5.1].

(٥-٤) انحدار عبر نقطة الأصل

يُعرف أحيانا أن خط الانحدار يمر عبر نقطة الأصل (0,0). ويحدث هذا، على سبيل المثال، عندما يكون X عدد الوحدات المنتجة وY التكلفة المتغيرة، ولذلك، فإن $Y = 0$ تعريفا عندما $X = 0$. ومثال آخر، حيث، X عدد أنواع الدخان المخزنة في سوق مركزية في تجربة (بما في ذلك الأسواق المركزية التي لاتخزن أية أنواع) وY حجم مبيعات الدخان في السوق المركزي. ونموذج الخطأ الطبيعي لهذه الحالات هو نموذج الانحدار (3.1) نفسه باستثناء أن $\beta_0 = 0$.

$$Y_i = \beta_1 X_i + \varepsilon_i \qquad (5.11)$$

حيث β_1 معلمة، X_i ثوابت معروفة، ε_i مستقلة $N(0, \sigma^2)$.

ودالة الانحدار للنموذج (5.11) هي:

$$E\{Y\} = \beta_1 X \qquad (5.12)$$

ويُحسب مقدر المربعات الدنيا لـ β_1 بجعل:

$$Q = \sum (Y_i - \beta_1 X_i)^2 \qquad (5.13)$$

أصغر مايمكن بالنسبة لـ β_1 والمعادلة الطبيعية الناتجة هي:

$$\sum X_i (Y_i - b_1 X_i) = 0 \qquad (5.14)$$

وتقود إلى المقدر النقطي

$$b_1 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad (5.15)$$

و b_1 المعطى في (5.15) هو أيضا مقدر الإمكانية العظمى.

وكمقدر غير منحاز لِ $E\{Y\}$ نجد:

$$\hat{Y} = b_1 X \qquad (5.16)$$

وتُعرَّف الرواسب، كالمعتاد، بالفرق بين القيم المشاهدة والتوفيقية:

$$e_i = Y_i - \hat{Y}_i = Y_i - b_1 X_i \qquad (5.17)$$

وكمقدر غير منحاز لـ σ^2 نجد:

$$MSE = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-1} = \frac{\sum (Y_i - b_1 X_i)^2}{n-1} = \frac{\sum e_i^2}{n-1} \qquad (5.18)$$

وسبب كون المقام n - 1 هو أننا فقدنا درجة حرية واحدة فقط عند تقدير المعلمة الوحيدة في دالة الانحدار (5.12).

جدول (٥-١) حدود ثقة لانحدار عبر الأصل

تقدير لـ		التباين المقدر	حدا الثقة
β_1	(5.19)	$s^2\{b_1\} = \frac{MSE}{\sum X_i^2}$	$b_1 \pm ts\{b_1\}$
$E\{Y_h\}$	(5.20)	$s^2\{\hat{Y}_h\} = \frac{X_h^2 MSE}{\sum X_i^2}$	$\hat{Y}_h \pm ts\{\hat{Y}_h\}$
$Y_{h(new)}$	(5.21)	$s^2\{Y_{h(new)}\} = MSE\left(1 + \frac{X_h^2}{\sum X_i^2}\right)$	$\hat{Y}_h \pm ts\{Y_{h(new)}\}$

حيث: $t = t(1 - \alpha/2, n - 1)$

وحدا الثقة لـ β_1 و$E\{Y_h\}$ ومشاهدة جديدة $Y_{h(new)}$ معروضة في الجدول (٥-١). لاحظ أن للمضاعف هنا درجة حرية، وهي درجات الحرية المرتبطة بـ MSE. وقد اشتقت النتائج في الجدول (٥-١) بطريقة مشابهة للنتائج السابقة الخاصة بنموذج الانحدار (3.1) وفي حين نواجه حدودا مثل $(X_i - \overline{X})^2$ أو $(X_h - \overline{X})^2$ في النموذج (3.1) مع جزء مقطوع نجد هنا X_i^2 و X_h^2 لأن نموذج الانحدار عبر الأصل.

مثال

تدير شركة تشارلز لتجهيزات أنابيب المياه 12 مستودعا. وفي محاولة لاختزال

إجراءات التخطيط والسيطرة، درس مستشار العلاقة بين وحدات العمل المنجزة X وتكلفة العمالة الكلية المتغيرة T في المستودعات خلال فترة الاختبار. والبيانات معطاة في العمودين (١) و(٢) من الجدول (٥-٢)، ويقدّم الشكل (٥-١) رسم انتشار المشاهدات.

استُخدم نموذج (5.11) للانحدار عبر الأصل حيث تتضمن Y التكاليف المتغيرة فقط وتبدو الشروط الأخرى لنموذج الانحدار مُحقَّقة كذلك. ومن العمودين (٣) و(٤) في الجدول (٥-٢) لدينا $\sum X_iY_i = 894{,}714$ و $\sum X_i^2 = 190{,}763$ وبالتالي:

$$b_1 = \frac{\sum X_iY_i}{\sum X_i^2} = \frac{894{,}714}{190{,}963} = 4.68527$$

ودالة الانحدار المقدَّرة:

$$\hat{Y} = 4.68527X$$

وفي الجدول (٥-٢) أُعطيت القيم التوفيقية في العمود (٥) والرواسب في العمود (٦) ورُسم خط الانحدار التوفيقي في الشكل (٥-١).

لتوضيح الاستقراءات في نموذج الانحدار عبر الأصل، افترض أننا نرغب تقدير فترة لـ β_1 بمعامل ثقة 95 بالمائة. بتربيع الرواسب في الجدول (٥-٢)، العمود (٦) وجمعها، نجد (الحسابات غير موضحة):

جدول (٥-٢) انحدار عبر الأصل ـ مثال المستودع

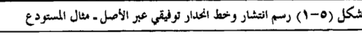
المستودع i	(١) وحدات العمل المنجزة X_i	(٢) تكلفة العمالة المتغيرة (بالدولار) Y_i	(٣) X_iY_i	(٤) X_i^2	(٥) $\hat{Y}_i$	(٦) e_i
1	20	114	2,280	400	93.71	20.29
2	196	921	180,516	38,416	918.31	2.69
3	115	560	64,400	13,225	538.81	21.19
4	50	245	12,250	2,500	234.26	10.74
5	122	575	70,150	14,884	571.60	3.40
6	100	475	47,500	10,000	468.53	6.47
7	33	138	4,554	1,089	154.61	−16.61
8	154	727	111,958	23,716	721.53	5.47
9	80	375	30,000	6,400	374.82	0.18
10	147	670	98,490	21,609	688.74	−18.74
11	182	828	150,696	33,124	852.72	−24.72
12	160	762	121,920	25,600	749.64	12.36
المجموع	1,359	6,390	894,714	190,963	6,367.28	22.72

$$MSE = \frac{\sum e_i^2}{n-1} = \frac{2,457.6}{11} = 223.42$$

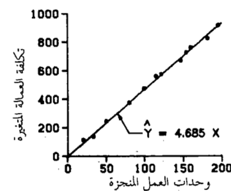
ومن الجدول (٥-٢) العمود (٤)، لدينا $\sum X_i^2 = 190,963$ وبالتالي:

$$s^2\{b_1\} = \frac{MSE}{\sum X_i^2} = \frac{223.42}{190,963} = 0.0011700 \qquad s\{b_1\} = 0.034205$$

ومن أجل 95 بالمائة معامل ثقة، نحتاج لِ $t(0.975; 11) = 2.201$ وحدا الثقة، وفقا لـ (5.19) في الجدول (٥-١)، هما $b_1 \pm ts\{b_1\}$ أو $4.68527 \pm 2.201(.034205)$



شكل (٥-١) رسم انتشار وخط انحدار توفيقي عبر الأصل - مثال المستودع



ولذلك فإن 95 بالمائة فترة ثقة لـ β_1 هي:

$$4.61 \le \beta_1 \le 4.76$$

وهكذا، بثقة 95 بالمائة، يقدَّر أن متوسط توزيع تكاليف العمالة الكلية المتغيرة يزداد بما يتراوح بين 4.61 و4.76 دولار، لكل وحدة عمل منجزة إضافية.

تعليقات

١- في الانحدار الخطّي عبر الأصل، لاتتوافر خاصية المربعات الدنيا من الشكل $\sum e_i = 0$ وبالتالي، لاتجمع الرواسب، عادة إلى الصفر، كما هو موضّح في الجدول (٥-٢)، العمود (٦) لمثال المستودع. وتأتي الخاصية الوحيدة للرواسب هنا من المعادلة الناظمية (5.14)، وتخصيصا، $\sum X_i e_i = 0$.

٢- في تقدير الفترة لـ $E\{Y_h\}$ أو التنبؤ بـ $Y_{h(new)}$ ، لاحظ أن الفترات في (5.20) و(5.21) في الجدول (٥-١) تتسع كلما ابتعد X_h عن الأصل. والسبب هو أن قيمة

دالة الانحدار معروفة بدقة عند نقطة الأصل، وبالتالي يزداد تأثير خطأ المعاينة في الميل b_1 أهمية كلما ابتعدت X_h عن الأصل.

٣- بما أنه ينبغي تقدير معلمة انحدار واحدة β_1 لدالة الانحدار (5.12)، فلا نحتاج هنا إلى طرق تقدير متزامنة لوضع عائلة من العبارات حول متوسطات استجابة متعددة. ومن أجل معامل ثقة α - 1 معطى، يمكن تطبيق العلاقة (5.20) بصورة متكررة مستخدمين نتائج العينة المعروفة لمستويات مختلفة لـ X وذلك لتوليد عائلة من العبارات والتي يبقى معامل ثقة العائلة من أجلها α - 1.

٤- ينبغي تقويم صلاحية نموذج الانحدار (5.11) شأنه شأن أي نموذج انحدار آخر. وحتى إذا عُرف أنه ينبغي لدالة الانحدار المرور من نقطة الأصل، فقد لاتكون الدالة خطّية أو قد لايكون تباين حدود الخطأ ثابتا. وفي الغالب لايمكن التأكد سلفا من أن خط الانحدار (3.1) يمر من نقطة الأصل وعندئذ يكون استخدام نموذج الانحدار (3.1) مع جزء مقطوع أكثر أمانا في الممارسة العملية. وإذا مر خط الانحدار من نقطة الأصل، فسوف يختلف b_0 عن 0 بخطأ معاينة صغير فقط، ومالم يكن حجم العينة صغيرا جدا، فليس لاستخدام نموذج الانحدار (3.1) مساوىء تذكر. وإذا لم يمر خط الانحدار من نقطة الأصل، فسيُجنّب استخدام نموذج الانحدار (3.1) صعوبات جدية ناتجة عن إجبار خط الانحدار على المرور من نقطة الأصل حيث لايكون ذلك مناسبا.

(٥-٥) تأثير أخطاء القياس

في مناقشتنا لنماذج الانحدار حتى الآن، لم نعتبر صراحة وجود أخطاء قياس في أي من X أو Y. والآن نفحص باختصار تأثير أخطاء القياس.

أخطاء قياس في Y

إذا وُجدت أخطاء قياس عشوائية في المتغير التابع Y، فلا تنشأ مشاكل جديدة عندما تكون هذه الأخطاء غير مرتبطة وغير منحازة (تنحو أخطاء القياس موجبها وسالبها إلى أن تلغي بعضها البعض). اعتبر، على سبيل المثال، دراسة العلاقة بين الوقت اللازم لإنهاء مهمة Y وتعقيد المهمة X. فقد لايقاس الوقت اللازم لإنهاء المهمة

بدقة لأن الشخص الذي يتولى ساعة القياس قد لايقوم بإيقافها بدقة في اللحظة المناسبة. ومادامت أخطاء القياس هذه ذات طبيعة عشوائية وغير مرتبطة وغير منحازة، فيمكن امتصاص أخطاء القياس هذه بسهولة في حد خطأ النموذج ε.

ويعكس حد خطأ النموذج التأثيرات المركّبة لعدد كبير من العوامل التي لم يشملها النموذج، وواحد منها ببساطة هو الأخطاء العشوائية التي تعود إلى عدم الدقة في عملية قياس Y.

أخطاء قياس في X

ولسوء الحظ، تنطبق حالة أخرى إذا عُرف المتغير المستقل X مع خطأ قياس. وعلى وجه التأكيد يكون X، في الغالب، معروفا بدون خطأ قياس، كأن يكون المتغير المستقل سعر منتج، أو عدد المتغيرات في مسألة أمثليّة، أو معدل الأجر لصنف من المستخدمين. وفي أحيان أُخر، على أي حال، قد تدخل أخطاء قياس في القيمة الملحوظة للمتغير المستقل، مثلا، عندما يكون ضغطا، أو درجة حرارة، أو سرعة خط إنتاج، أو عمر شخص.

وسنستخدم التوضيح الأخير في تطويرنا لطبيعة المسألة. افترض أننا نحدر دخل العامل، على أساس القطعة، على عمره. وليكن X_i العمر الحقيقي للمستخدم i و X_i^* العمرالذي أعطاه المستخدم في سجل المستخدمين. ولاحاجة للقول، إن العمرين لايتطابقان دائما. ونعرّف خطأ القياس δ_i كالتالي:

$$\delta_i = X_i^* - X_i \tag{5.22}$$

ونموذج الانحدار الذي نرغب دراسته هو:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{5.23}$$

ولكن، وبما أننا نلحظ X_i^* فقط، فيصبح النموذج (5.23):

$$Y_i = \beta_0 + \beta_1 (X_i^* - \delta_i) + \varepsilon_i \tag{5.24}$$

حيث استفدنا من (5.22) عند التعويض عن X_i. ويمكننا إعادة كتابة (5.24) كالتالي:

$$Y_i = \beta_0 + \beta_1 X_i^* + (\varepsilon_i - \beta_1 \delta_i) \tag{5.25}$$

وقد يبدو النموذج (5.25) وكأنه نموذج انحدار اعتيـادي بمتغير مستقل X وحد خطأ $\varepsilon\text{-}\beta_1\delta$ ، إلا أنه ليس كذلك. فالمتغير المستقل الملحوظ X_i^* هو متغير عشوائي مرتبط، كما سنرى، مع حد الخطأ $\varepsilon_i - \beta_1\delta_i$. وتتطلب النظرية (3.40) الخاصة بمتغيرات مستقلة عشوائية، أن يكون حد الخطأ مستقلا عن المتغير المستقل. وبالتالي لاتنطبق نتائج الانحدار كما نعرفها. على النموذج (5.25).

ونجد، بالبداهة أن $\varepsilon_i\text{-}\beta_1\delta_i$ غير مستقلة عن X_i^* حيث تقيّد العلاقة (5.22) $X_i^* - \delta_i$ لتكون مساوية لـ X_i. ولتحديد التبعية تحديدا رسميا، دعنا نفترض أن:

$$E\{\delta_i\} = 0 \qquad (5.26a)$$
$$E\{\varepsilon_i\} = 0 \qquad (5.26b)$$
$$E\{\delta_i\varepsilon_i\} = 0 \qquad (5.26c)$$

لاحظ أن (5.26a) تتضمن $E\{X_i^*\} = E\{X_i + \delta_i\} = X_i$ وتفترض (5.26c) أن خطأ القياس δ_i لايرتبط مع خطأ النموذج ε_i إذ لدينا من (1.21a) $\sigma\{\delta_i,\varepsilon_i\} = E\{\delta_i\varepsilon_i\}$ باعتبار أن $E\{\delta_i\}=E\{\varepsilon_i\}=0$ كما افترضنا في (5.26a) و(5.26b).

ونرغب الآن إيجاد التغاير:

$$\sigma\{X_i^*, \varepsilon_i - \beta_1\delta_i\} = E\{[X_i^* - E\{X_i^*\}][(\varepsilon_i - \beta_1\delta_i) - E\{\varepsilon_i - \beta_1\delta_i\}]\}$$
$$= E\{(X_i^* - X_i)(\varepsilon_i - \beta_1\delta_i)\}$$
$$= E\{\delta_i(\varepsilon_i - \beta_1\delta_i)\}$$
$$= E\{\delta_i\varepsilon_i - \beta_1\delta_i^2\}$$

ومن (5.26c) لدينا الآن $E\{\delta_i\varepsilon_i\} = 0$ ومن (1.15a) لدينا $E\{\delta_i^2\} = \sigma^2\{\delta_i\}$ باعتبار أننا افترضنا $E\{\delta_i\} = 0$ في (5.26a) ولذلك نجد:

$$\sigma\{X_i^*, \varepsilon_i - \beta_1\delta_i\} = -\beta_1\sigma^2\{\delta_i\} \qquad (5.27)$$

وهذا التغاير ليس صفرا إذا كانت هناك علاقة انحدار خطّية بين X وY.

وإذا طُبقت طرق المربعات الدنيا القياسية على النموذج (5.25)، فالمقدّران b_o وb_1 منحازان، وكذلك يفتقران إلى خاصية الاتساق. وتواجهنا مصاعب جمّة في تطوير مقدّرين غير منحازين عندما توجد أخطاء قياسات في X، وإحدى الطرق هي فرض شروط قاسية على المسألة ـ على سبيل المثال، وضع افتراضات قوية إلى حد ما حول

خواص توزيعات δ_i وتغاير مع δ_i وهلم جرا. وطريقة أخرى تتمثل في استخدام متغيرات إضافية نعلم أنها على صلة بقيم X الحقيقية، ولكنها ليست مرتبطة مع أخطاء القياس δ. تسمى مثل هذه المتغيرات *متغيرات أدواتية* لأنها تُستخدم كأداة في دراسة العلاقة بين X و Y. والمتغيرات الأدواتية تجعل من الممكن إيجاد مقدّرات متّسقة لمعالم الانحدار.

ويمكن العثور على مناقشات للطرق الممكنة ومراجع إضافية في كتب متخصصة مثل مرجع [5.2].

ملاحظة

قد نتساءل لماذا التمييز بين الحالة التي يكون X فيها متغيرا عشوائيا، والتي درسناها في الفصل الثالث، وبين الحالة التي تخضع فيها X لأخطاء قياس عشوائية، ولماذا توجد مشاكل خاصة بالحالة الأخيرة. عندما يكون X متغيرا عشوائيا، فإنه لايخضع لسيطرة المحلل وسيتغير عشوائيا من محاولة إلى أخرى، كأن يكون X عدد الأشخاص الداخلين إلى متجر في يوم ما. وإذا لم يكن هذا المتغير X خاضعا لخطأ قياس، على أي حال، فيمكن التحقق بدقة من قيمته في محاولة معينة. وهكذا، إذا لم توجد أخطاء قياس في تعداد الأشخاص الداخلين إلى متجر في يوم ما، فلدى المحلل معلومات دقيقة لدراسة العلاقة بين عدد الأشخاص الداخلين إلى المتجر والمبيعات، هذا بالرغم من أن مستويات عدد الأشخاص الداخلين بالفعل إلى المتجر لايخضع للسيطرة. وعلى الوجه الآخر، إذا كانت أخطاء القياس في عدد الأشخاص الداخلين إلى المتجر موجودة فسنحصل على صورة مشوهة للعلاقة بين عدد الأشخاص والمبيعات ذلك لأنه سيجري توفيق المبيعات الملحوظة لأعداد غير صحيحة من الأشخاص. وهذا التأثير التشويهي لأخطاء القياس موجود سواء أكان X مثبتا أم عشوائيا.

نموذج بيركسون

توجد هناك حالة لاتشكل معها أخطاء القياس في X أي مشكلة. لوحظت هذه الحالة أولا من قبل بيركسون [مرجع 5.3]. فكثيرا مايوضع المتغير المستقل في تجربة عند قيمة مستهدفة. فمثلا، في تجربة تتناول تأثير درجة الحرارة في إنتاجية ضارب آلة

كاتبة، قد توضع درجة الحرارة عند مستويات مستهدفة مثل 68 °F و 70 °F وما شابه، وذلك بالسيطرة على درجة الحرارة من خلال ثرموستات. فدرجة الحرارة الملحوظة X_i مثبتة هنا، في حين أن درجة الحرارة الفعلية متغير عشوائي باعتبار أن الثرموستات قد لايكون دقيقا تماما. وهناك حالة مشابهة عندما يوضع ضغط الماء تبعا لمنظم أو عندما نختار للدراسة عمالا من عمر محدد طبقا لسجلات عملهم.

وفي جميع هذه الحالات، نجد أن المشاهدة X_i^* كمية مثبتة بينما القيمة الصحيحة غير الملحوظة X_i هي متغير عشوائي. وخطأ القياس، كما سبق:

$$\delta_i = X_i^* - X_i \tag{5.28}$$

وعلى أي حال، لاتوجد هنا قيود على العلاقة بين X_i^* و δ_i لأن X_i^* كمية مثبتة. ومرة ثانية نفترض أن $E\{\delta_i\} = 0$.

ولايزال من الممكن تطبيق النموذج (5.25)، الذي نحصل عليه عندما نستبدل X_i بـ $X_i^* - \delta_i$ في حالة بيركسون:

$$Y_i = \beta_0 + \beta_1 X_i^* + (\varepsilon_i - \beta_1 \delta_i) \tag{5.29}$$

والقيمة المتوقعة لحد الخطأ $E\{\varepsilon_i - \beta_1\delta_i\}$ تساوي الصفر كالسابق، لأن $E\{\varepsilon_i\} = 0$ و $E\{\delta_i\} = 0$ وبالإضافة إلى ذلك فإن $\varepsilon_i - \beta_1\delta_i$ غير مرتبطة مع X_i^* الآن، لأن X_i^* ثابت في حالة بيركسون. وبالتالي تتحقق الشروط التالية لنموذج الانحدار العادي:

١- توقع حدود الخطأ يساوي الصفر.

٢- المتغير المستقل ثابت، وبالتالي لاترتبط به حدود الخطأ.

وهكذا يمكن تطبيق طرق المربعات الدنيا على حالة بيركسون بدون تعديل، وسيكون المقدّران b_0 و b_1 غير منحازين. وإذا أمكن كالمعتاد افتراض طبيعية الأخطاء $\varepsilon_i - \beta_1\delta_i$ وثبات تباينها فيمكن الإستفادة من تقديرات الفترة والاختبارات المعتادة.

(٥-٦) تنبؤات عكسية

أحيانا، يُستخدم نموذج انحدار Y على X للتنبؤ بقيمة X التي نشأت عنها مشاهدة جديدة، ويعرف هذا بالتنبؤ العكسي. ونوضّح التنبؤات العكسية بمثالين:

١- حَدر محلل رابطة تجارية سعر بيع منتج Y على تكلفته X وذلك في الشركات الـ 15 الأعضاء في الرابطة. سعر البيع $Y_{h(new)}$ لشركة أخرى لاتنتمي للرابطة التجارية معروف، ومن المرغوب تقدير التكلفة $X_{h(new)}$ لهذه الشركة.

٢- أُجري تحليل انحدار لنقص مستوى الكوليسترول Y المقابل لجرعة دواء جديد X وذلك بالاستناد إلى مشاهدات من 50 مريضا. ويعالج طبيب مريضا جديدا ينبغي له إنقاص مستوى الكوليسترول بمقدار $Y_{h(new)}$ ويرغب في تقدير مستوى الجرعة X المناسب للوصول إلى النقص المطلوب $Y_{h(new)}$ في مستوى الكوليسترول.

في التنبؤ العكسي، نفترض كالسابق نموذج الانحدار (3.1):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (5.30)$$

ونحصل كالمعتاد على دالة الانحدار المقدرة بالاستناد إلى n مشاهدة:

$$\hat{Y} = b_0 + b_1 X \qquad (5.31)$$

ومع توافر مشاهدة جديدة $Y_{h(new)}$ نرغب في تقدير المستوى $X_{h(new)}$ الذي نشأت عنه هذه المشاهدة الجديدة. ونحصل على تقدير نقطي عادي بحل (5.31) من أجل X علما أن $Y_{h(new)}$ معروفة:

$$\hat{X}_{h(new)} = \frac{Y_{h(new)} - b_0}{b_1} \qquad b_1 \neq 0 \qquad (5.32)$$

حيث يشير $\hat{X}_{h(new)}$ إلى المقدِّر النقطي للمستوى الجديد $X_{h(new)}$ ويحتوي الشكل (٥-٢) على تمثيل لهذا المقدر النقطي في مثال سنناقشه بعد قليل. وفي الواقع إن $\hat{X}_{h(new)}$ هو مقدِّر الإمكانية العظمى لـ $X_{h(new)}$ في نموذج الانحدار (3.1).

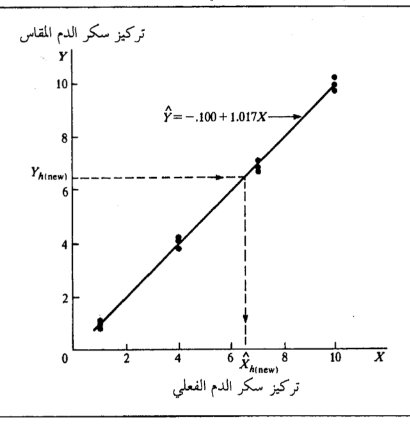
ويمكن تبيان أن الـ $(1 - \alpha)$ حدي ثقة لـ $X_{h(new)}$ هما:

$$\hat{X}_{h(new)} \pm t(1-\alpha/2; n-2) s\{\hat{X}_{h(new)}\} \qquad (5.33)$$

حيث:

$$s^2\{\hat{X}_{h(new)}\} = \frac{MSE}{b_1^2}\left[1 + \frac{1}{n} + \frac{(\hat{X}_{h(new)} - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right] \tag{5.33a}$$

شكل (٥-٢) رسم انتشار وخط الانحدار التوفيقي - مثال المعايرة.



مثال

في إحدى التجارب استخدم باحث في علوم الطب طريقة جديدة لقياس التركيز المنخفض لسكر الدم (Y) وذلك في عينة، تركيز السكر فيها معروف (X). وفي مجمل التجربة استُخدمت أربعة مستويات للتركيز. ولدى توفيق نموذج الانحدار (3.1) وجدنا النتائج التالية:

$n = 12$ $\quad b_0 = 0.100$ $\quad b_1 = 1.017$ $\quad MSE = .0272$

$s\{b_1\} = .0142$ $\quad \overline{X} = 5.500$ $\quad \overline{Y} = 5.492$ $\quad \sum(X_i - \overline{X})^2 = 135$

$$\hat{Y} = -.100 + 1.017X$$

رُسمت البيانات وخط الانحدار المقدَّر في شكل (٥-٢).

يرغب الباحث أولا التأكد من أن هناك صلة خطّية بين المتغيرين.وقد قام باختبار $H_0 : \beta_1 = 0$ ضد $H_a : \beta_1 \neq 0$ مستفيدا من إحصاءة الاختبار $t^* = b_1 / s\{b_1\} = 1.017 / 0.0142 = 71.6$ ومستخدما $\alpha = 0.05$ وحيث إن $t(0.975; 10) = 2.228$ و$|t^*| = 71.6 > 2.228$ فقد استنتج أن $\beta_1 \neq 0$ أي أنه توجد صلة خطّية بين التركيز المقاس والتركيز الفعلي.

ويرغب الباحث الآن استخدام علاقة الانحدار لمريض جديد أعطت طريقة القياس الجديدة من أجله $Y_{h(new)} = 6.52$ ويريد تقدير التركيز الفعلي $X_{h(new)}$ لهذا المريض مستخدما 95 بالمائة فترة ثقة.

باستخدام (5.32) و ($5.33a$) نجد:

$$\hat{X}_{h(new)} = \frac{6.52 - (-0.100)}{1.017} = 6.509$$

$$s^2\{\hat{X}_{h(new)}\} = \frac{0.0272}{(1.017)^2}\left[1 + \frac{1}{2} + \frac{(6.509 - 5.500)^2}{135}\right] = 0.0287$$

وبالتالي $s\{\hat{X}_{h(new)}\} = 0.1694$ ، ونحتاج إلى $t(.975;\ 10) = 2.228$ وباستخدام (5.33) نجد حدي الثقة $6.509 \pm 2.228\ (1.694)$. وبالتالي تكون الـ 95 بالمائة فترة ثقة:

$$6.13 \leq X_{h(new)} \leq 6.89$$

وهكذا يمكن الاستنتاج بـ 95 بالمائة ثقة أن تركيز السكر الفعلي يقع بين 6.13 و6.89 وهو على وجه التقريب ±6 بالمائة خطأ، وقد اعتبره الباحث خطأ معقولا.

تعليقات

١- تُعرف مسألة التنبؤ المعكوس، كذلك، ***بمسألة المعايرة***، إذ تُطبق عندما تكون قياسات تقريبية وغير مكلفة وسريعة لـ Y على صلة بقياسات X محكمة، ومُكلفة في الغالب، وتستغرق زمنا، وتستند إلى n مشاهدة. ويُستخدم نموذج الانحدار الناتج لتقدير القياس الدقيق $X_{h(new)}$ المقابل لقياس تقريبي جديد $Y_{h(new)}$ وأوضحنا هذا الاستخدام في مثال المعايرة.

٢- تكون فترة الثقة التقريبية (5.33) مناسبة. إذا كانت الكمية:

$$\frac{\left[t(1-\alpha/2;n-2)\right]^2\ MSE}{b_1^2\ \Sigma(X_i-\overline{X})^2} \tag{5.34}$$

صغيرة، ولنقل أقل من 0.1 وهذه الكمية هي، في مثال المعايرة:

$$\frac{(2.228)^2\,(0.0272)}{(1.017)^2\,(135)}=0.00097$$

وبالتالي تكون فترة الثقة التقريبية مناسبة هنا.

٣- نحصل بسهولة على فترات تنبؤ متزامنة قائمة على عدد g من القياسات الملحوظة الجديدة والمختلفة $X_{h(new)}$ بـ $(1 - \alpha)$ معامل ثقة عائلي، مستخدمين إما طريقة بونفرّوني أو طريقة شفّيه اللتين ناقشناهما في الفقرة (٥ ـ ٣). ونستبدل: $B = t(1 - \alpha/2g;\ n - 2)$ أو $S = [gF(1-\alpha;g,n-2)]^{1/2}$ بالقيمة $t(1-\alpha/2;n-2)$ في (5.33).

(٥-٧) اختيار مستويات X

عند الحصول على بيانات الانحدار في تجربة، فإن مستويات X التي ستقاس من أجلها المشاهدات Y تكون تحت سيطرة المجرب. ومن بين أشياء أخرى، على المجرب أن يأخذها في اعتباره:

١- كم مستو من مستويات X ينبغي أن تتناوله الدراسة؟

٢- ماهما المستويان المتطرفان (أصغر مستوى وأكبر مستوى)؟

٣- كيف ستكون مسافات بقية المستويات بعضها عن بعض، إن وُجدت؟

٤- ماعدد المشاهدات التي ينبغي أخذها عند كل مستوى من مستويات X؟

لاتوجد إجابة بمفردها على هذه الأسئلة، لأن الأجوبة تختلف باختلاف الغرض من تحليل الانحدار. وكما ذكرنا سابقا فإن الأهداف الممكنة لتحليل الانحدار هي أهداف متنوعة. وقد يكون الهدف الرئيس تقدير ميل خط الانحدار، أو في بعض الحالات تقدير الجزء المقطوع. وفي العديد من الحالات يكون الهدف الرئيس هو التنبؤ بمشاهدة جديدة أو أكثر، أو تقدير متوسط استجابة أو أكثر. وعندما تكون دالة الانحدار منحنية، فقد يكون الهدف الرئيس تحديد أعلى أو أدنى متوسط استجابة. ويبقى أن نقول إن الغرض الرئيس، في بعض الأحيان هو تحديد طبيعة دالة الانحدار.

• ولتوضيح كيفية تأثيرات الهدف في التصميم، لنتأمل تباينات b_0 ، b_1 و $Y_{h(new)}$ والتي استُنبطت سابقا لنموذج الانحدار (3.1):

$$\sigma^2\{b_0\} = \sigma^2 \frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2} = \sigma^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum(X_i - \overline{X})^2}\right] \quad (5.35)$$

$$\sigma^2\{b_1\} = \frac{\sigma^2}{\sum(X_i - \overline{X})^2} \quad (5.36)$$

$$\sigma^2\{\hat{Y}_h\} = \sigma^2\left[\frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right] \quad (5.37)$$

$$\sigma^2\{Y_{h(new)}\} = \sigma^2\left[1 + \frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right] \quad (5.38)$$

فتباين الميل يكون أقل مايمكن إذا كان $\sum(X_i - \overline{X})^2$ أكبر مايمكن، ويمكن إنجاز ذلك باستخدام مستويين لـ X، عند بداية ونهاية مدى النموذج، ووضع نصف المشاهدات عند كل من المستويين. وبالطبع إذا لم يكن الباحث متأكدا من خطّية دالة الانحدار، فسوف يتردد في استخدام مستويين فقط إذ أن النقطتين لاتوفران أية معلومات عن الانحرافات الممكنة عن الخطّية.

وإذا كان الهدف الرئيس هو تقدير β_0، فلا أهمية لعدد المستويات أو لمواقعها، ما دامت $\overline{X} = 0$. وعلى الوجه الآخر، عند تقدير متوسط الاستجابة المقابل لـ X أو التنبؤ بمشاهدة جديدة عند X_h، فمن الأفضل استخدام مستويات X بحيث يكون $\overline{X} = X_h$. وإذا أردنا تقدير عدة مستويات استجابة أو التنبؤ بعدة مشاهدات جديدة، فمن الأفضل نشر مستويات X بحيث يكون $\overline{X}$ في مركز مستويات X_h التي هي موضع الاهتمام.

وبالرغم من أن عدد ومستويات X ومسافاتها بعضها عن بعض تعتمد بقوة على الغرض الرئيس من تحليل الانحدار، يمكن إعطاء بعض النصائح العامة، لاستخدامها كنقطة انطلاق، على الأقل ويقترح د. ر. كوكس مايلي:

استخدم مستويين عندما يكون الهدف بصورة رئيسة، فحص ما إذا كان... (للمتغير المستقل)... تأثير، وفي أي اتجاه يعمل هذا التأثير. واستخدم ثلاثة مستويات عندما تتوقع أن يكون وصف منحنى الاستجابة من خلال ميله وانحنائه وصفا مناسبا؛ وينبغي أن يغطي هذا

معظم الحالات. واستخدم أربعة مستويات إذا كان من المهم القيام بمزيد من الفحوص لشكل منحني الاستجابة واستخدم أكثر من أربعة مستويات عندما يكون المطلوب تقدير الشكل التفصيلي لمنحنى الاستجابة، أو عندما نتوقع أن يرتفع المنحنى إلى قيمة مقاربة، أو، بصورة عامة، لتبيان نواح لايمكن وصفها وصفا مناسبا من خلال الميل و الانحناء. وفيما عدا الحالات الأخيرة هذه يكون من المستحسن عامة استخدام مستويات متساوية البعد بعضها عن بعض وبأعداد متساوية من المشاهدات عند كل مستوى [مرجع 5.4].

مراجع

[5.1] Miller, R. G., Jr. *Simultaneous Statistical Inference*; 2nd ed. New York: Springer Verlag, 1981, pp. 114 - 16.

[5.2] Fuller, W. A. *Measurement Error Models*. New York: John Wiley & Sons, 1987.

[5.3] Berkson, J. "Are There Two Regressions?" *Journal of the American Statistical Association*, 45 (1950), pp. 164 - 80.

[5.4] Cox, D. R. *Planning of Experiments*. New York: John Wiley & Sons, 1958, 141 - 42.

مسائل

(٥-١) عند تطوير فترات ثقة مشتركة لـ β_0 و β_1 بطريقة بونفرّوني بـ 90 في المائة معامل ثقة عائلي، هل يتضمن هذا أن فترات الثقة لـ β_0 ستكون غير صحيحة في 10 بالمائة من المرات؟ إن فترات الثقة لـ β_0 غير صحيحة في 5 بالمائة من المرات وأن فترات الثقة لـ β_1 ستكون غير صحيحة في 5 بالمائة من المرات؟ ناقش.

(٥-٢) بالعودة إلى المسألة (٣-١) افترض أن الطالب ضمَّ فترتي الثقة في مجموعة ثقة واحدة. ماذا يمكنك القول عن معامل الثقة العائلي لهذه المجموعة ؟

(٥-٣) بالعودة إلى مسألتي **صيانة الحاسبات** (٢-١٨) و(٣-٥).

ا ـ هل يميل b_0 و b_1 هنا إلى الخطأ في الاتجاه نفسه أم في اتجاهين متضادين؟ وضح.

ب ـ أوجد فترتي ثقة بونفرّوني المشتركة لـ β_0 و β_1 مستخدما 95 بالمائة معامل ثقة عائلي.

جـ ـ يقترح مستشار أن β_0 ينبغي أن يكون صفرا وأن β_1 ينبغي أن يساوي 14.0. هل تدعم فترتا ثقتك المشتركة في (ب) وجهة النظر هذه؟

(٥ـ٤) بالعودة إلى مسألة **تكسر الشحنات** (٢-١٩).

ا ـ هل تميل b_0 و b_1 هنا إلى الخطأ في الاتجاه نفسه أم في اتجاهين متضادين؟ وضح.

ب ـ أوجد فترتي ثقة بونفرّوني المشتركة لـ β_0 و β_1 مستخدما 99 بالمائة معامل ثقة عائلي. فسّر فترتي ثقتك.

(٥ـ٥) بالعودة إلى مسألة **صلابة البلاستيك** (٢-٢٠).

ا ـ أوجد فترتي ثقة بونفرّوني المشتركة لـ β_0 و β_1 مستخدما 90 بالمائة معامل ثقة عائلي. فسّر فترتي ثقتك.

ب ـ هل b_0 و b_1 مرتبطان إيجابا أم سلبا هنا؟ هل ينعكس هذا في فترتي ثقتك المشتركة في الجزء (ا)؟

جـ ـ ماهو معنى معامل ثقة العائلة في الجزء (ا)؟

(٥ـ٦) عُد إلى مسألة **كتلة العضلة** (٢-٢٥).

ا ـ أوجد فترتي ثقة بونفرّوني المشتركة لـ β_0 و β_1 مستخدما 99 بالمائة معامل ثقة عائلي. فسّر فترتي ثقتك.

ب ـ هل تميل b_0 و b_1 هنا إلى الخطأ في الاتجاه نفسه أم في اتجاهين متضادين؟ وضّح.

جـ ـ يقترح باحث أن ينبغي أن β_0 تساوي تقريبا 160 وأن β_1 ينبغي أن تقع بين 1.9- و 1.5- . هل تدعم فترتا ثقتك المشتركة في الجزء (أ) هذا التوقع؟

(٥ـ٧) بالعودة إلى مسألتي **صيانة الحاسبات** (٢-١٨) و(٣-٥).

ا ـ قدّر توقع عدد الدقائق المستغرقة عندما يوجد 3، 5 و 7 آلات، على الترتيب، تحتاج إلى خدمة. استخدم تقديرات فترة بـ 90 بالمائة معامل ثقة عائلي معتمدا على طريقة ووركنج ـ هوتلّنج.

ب ـ تم ترتيب ندائي خدمة صيانة وقائية، وعدد الآلات التي ستجري خدمتها

فيهما 4 و 7 على الترتيب. ونرغب في عائلة من فترتي تنبؤ للزمنين اللذين سيستغرقهما النداءان بـ 90 بالمائة معامل ثقة عائلي. أي من الطريقتين، شفّيه أم بونفرّوني، ستقدم حدي تنبؤ أضيق هنا؟.

جـ ـ أوجد العائلة من حدي التنبؤ المطلوبة في الجزء (ب) مستخدما الطريقة الأكثر كفاءة.

(٥ـ٨) بالعودة إلى مسألة **تكسر الشحنات** (٢-١٩).

ا ـ من المرغوب فيه إيجاد تقديرات فترة لمتوسط عدد الأنبولات المكسورة عندما يوجد 0 ، 1 و2 من التحويلات للشحنة، وذلك باستخدام 95 بالمائة معامل ثقة عائلي. أوجد فترات الثقة المرغوبة مستخدما طريقة ووركنج – هوتلّنج.

ب ـ هل فترات الثقة التي حصلت عليها في الجزء أكثر كفاءة من فترات (ا) بونفرّوني هنا؟ وضح.

جـ ـ ستتعرض الشحنات الثلاث القادمة إلى 0، 1 و 2 تحويلا على الترتيب. أوجد حدي تنبؤ لعدد الأنبولات المكسورة لكل من هذه الشحنات الثلاث مستخدما طريقة شفّيه و 95 بالمائة معامل ثقة عائلي.

د ـ هل ستكون طريقة بونفرّوني أكثر كفاءة في تطوير فترات التنبؤ في الجزء (جـ)؟ وضّح.

(٥ـ٩) بالعودة إلى مسألة **صلابة البلاستيك** (٢ـ٢٠).

ا ـ ترغب الإدارة في إيجاد تقديرات فترة لمتوسطات الصلابة عندما يكون الوقت المنصرم 20، 30 و40 ساعة، على الترتيب، احسب فترات الثقة المرغوبة مستخدما طريقة بونفرّوني و 90 بالمائة معامل ثقة عائلة. ما هو معنى معامل ثقة العائلة هنا؟.

ب ـ هل طريقة بونفرّوني المستخدمة في الجزء (ا) الطريقة الأكثر كفاءة التي يمكن استخدامها هنا؟ وضّح.

جـ ـ ستُقاس الوحدتان القادمتان من أجل وقت منصرم مقداره 30 و 40 ساعة على الترتيب. تنبأ بصلابة كل من هاتين الوحدتين مستخدما الطريقة الأكثر كفاءة مع 90 بالمائة معامل ثقة عائلة.

(٥-١٠) بالعودة إلى مسألة **كتلة العضلة** (٢- ٢٥).

ا ـ يهتم اختصاصي تغذية، على وجه الخصوص، بمتوسط كتلة العضلة لنساء أعمارهن 45، 55 و65. أوجد فترات ثقة مشتركة للمتوسطات موضع الاهتمام مستخدما طريقة ووركنج ـ هوتلّنج، و90 بالمائة معامل ثقة عائلة.

ب ـ هل طريقة ووركنج ـ هوتلّنج هي الطريقة الأكثر كفاءة التي يمكن استخدامها في الجزء (ا)؟ وضّح.

جـ ـ اتصلت ثلاث نساء أخريات أعمارهن 48، 59 و74 سنة باختصاصي التغذية. تنبأ بكتلة العضلة لكل من هؤلاء النسوة الثلاث مستخدما طريقة بونفرّوني و 95 بالمائة معامل ثقة عائلة.

د ـ لاحقا، يرغب اختصاصي التغذية التنبؤ بكتلة العضلة لامرأة رابعة عمرها 64، مع 95 بالمائة معامل ثقة عائلة للتنبؤات الأربعة. هل يجب إعادة حساب فترات التنبؤ الثلاث في الجزء (جـ)؟

هل يكون جوابك هذا صحيحا أيضا لو استخدمت طريقة شفّيه لوضع فترات التنبؤ؟

(٥-١١) صرح عالم سلوكي حديثا: "لست أبدا متأكدا مما إذا كان خط الانحدار يمر من نقطة الأصل. وبالتالي فسوف لا أستخدم مثل هذا النموذج "علّق.

(٥-١٢) **أخطاء مطبعية**. فيما يلي عدد لوحات الطباعة لمخطوطة (X) والتكلفة الكلية بالدولار لتصحيح الأخطاء المطبعية (Y) وذلك في عينة عشوائية من الطلبات الحديثة التي تعهدتها شركة متخصصة في مخطوطات تقنية. وبما أن Y ينطوي على متغير تكاليف فقط فقد رغب محلل في تحديد ما إذا كان نموذج الانحدار عبر نقطة الأصل (5.11) ملائما لدراسة العلاقة بين المتغيرين:

12	11	10	9	8	7	6	5	4	3	2	1	i
6	4	10	18	25	30	25	14	10	10	12	7	X_i
107	75	177	324	457	540	446	250	178	191	213	128	Y_i

ا ـ قُم بتوفيق نموذج الانحدار (5.11) واعرض دالة الانحدار المقدَّرة.

ب ـ ارسم دالة الانحدار المقدَّرة والبيانات. هل يبدو أن دالة انحدار خطِّية مارة من نقطة الأصل تشكل توفيقا جيدا هنا ؟ علّق.

جـ ـ عند تقدير تكاليف معالجة طلبيات محتملة. استخدمت الإدارة معيارا من 17.50 دولارا لكل لوحة كتكلفة تصحيح أخطاء مطبعية. اختبر ما إذا كان ينبغي تحسين هذا المعيار أم لا؛ استخدم $\alpha = 0.02$. اعرض البديلين، وقاعدة القرار والنتيجة.

د ـ أوجد فترة تنبؤ لتكلفة التصحيح لعمل قادم يتضمن 10 لوحات. استخدم 98 بالمائة معامل ثقة.

(٥ـ١٣) بالعودة إلى مسألة أخطاء مطبعية (٥ـ١٢).

ا ـ أوجد الرواسب e_i. هل تجمع إلى الصفر ؟ ارسم الرواسب مقابل القيم التوفيقية $\hat{Y}_i$. ماهي النتائج التي يمكن استخلاصها من رسمك ؟

ب ـ قم باختبار رسمي لنقص التوفيق في انحدار خطّي عبر الأصل ; استخدم $\alpha = 0.01$. اعرض البديلين وقاعدة القرار والنتيجة. ماهي القيمة -P للاختبار؟.

(٥ـ١٤) بالعودة إلى مسألة **المعدل التراكمي** (٢ـ١٧). افترض أن نموذج الانحدار الخطي عبر الأصل (5.11) مناسب.

ا ـ قم بتوفيق نموذج الانحدار (5.11) واعرض دالة الانحدار المقدَّرة.

ب ـ قدّر β_1 بـ 95 بالمائة فترة ثقة.فسر تقديرك بفترة.

جـ ـ قدّر المعدل التراكمي المتوسط للطلاب المستجدين الذين نالوا درجة 5.7 في امتحان الدخول. استخدم 95 بالمائة فترة ثقة.

(٥ـ١٥) بالعودة إلى مسألة **المعدل التراكمي** (٥ـ١٤).

ا ـ ارسم خط الانحدار التوفيقي والبيانات، هل تبدو دالة الانحدار الخطّية عبر الأصل توفيقا جيدا هنا.

ب ـ أوجد الرواسب e_i. هل تجمع إلى الصفر؟ ارسم الرواسب مقابل القيم التوفيقية $\hat{Y}_i$. ماهي النتائج التي يمكن استخلاصها من رسمك ؟

جـ ـ قم باختبار رسمي لنقص التوفيق لانحدار خطّي عبر الأصل ؛ استخدم $\alpha = 0.005$ اعرض البديلين وقاعدة القرار والنتيجة. ما هي القيمة ـP للاختبار؟

(٥ـ١٦) بالعودة إلى مسألة **صيانة الحاسبات** (٢–١٨). افترض أن نموذج الانحدار الخطّي عبر الأصل (5.11) مناسب.

ا ـ أوجد دالة الانحدار المقدَّرة.

ب ـ قدر β_1 بـ 90 بالمائة فترة ثقة. فسّر تقديرك بفترة.

جـ ـ تنبأ بزمن الخدمة في نداء جديد تتم فيه خدمة ست آلات. استخدم 90 بالمائة فترة ثقة.

(٥ـ١٧) بالعودة إلى مسألة **صيانة الحاسبات** (٥–١٦)

ا ـ ارسم خط الانحدار التوفيقي والبيانات. هل يبدو أن دالة انحدار خطّي عبر الأصل تشكّل توفيقا جيدا هنا؟

ب ـ أوجد الرواسب e_i. هل تجمع إلى الصفر؟ ارسم الرواسب مقابل القيم التوفيقية $\hat{Y}_i$. ما هي النتائج التي يمكن استخلاصها من رسمك؟.

جـ ـ قم باختبار رسمي لنقص التوفيق لخط انحدار عبر الأصل؟ استخدم اعرض البديلين وقاعدة القرار والنتيجة. ماهي القيمة ـ P للاختبار ؟.

(٥ـ١٨) بالعودة إلى مسألة **صلابة البلاستيك** (٢–٢٠). افترض أن الأخطاء تظهر في X لأنه طُلب من فني المختبر أن يقيس صلابة العينة (Y_i) الموافقة لوقت منصرم محدد سلفا (X_i) ولكن التوقيت غير دقيق، ولهذا يتغير الوقت الحقيقي المنصرم عشوائيا عن الوقت المنصرم المحدد سلفا. هل سيكون مقدر المربعات الدنيا هنا منحازا؟ ناقش.

(٥ـ١٩) بالعودة إلى مسألتي **المعدل التراكمي** (٢–١٧) و(٣–٤). حصل طالب جديد على معدل تراكمي 3.4 في السنة الأولى.

ا - أوجد 90 بالمائة فترة ثقة لدرجة الطالب في امتحان الدخول. فسّر فترة ثقتك.

ب - هل تحققت شروط المعيار (5.34) فيما يتعلق بصلاحية فترة الثقة التقريبية؟

(٥-٢٠) بالعودة إلى مسألة **صلابة البلاستيك** (٢-٢٠). أظهر القياس لوحدة اختبار جديدة 238 وحدة برنل للصلابة.

ا - أوجد 99 بالمائة فترة ثقة للزمن المنصرم قبل قياس الصلابة. فسّر فترة ثقتك.

ب - هل تحققت شروط المعيار (5.34) فيما يتعلق بصلاحية فترة الثقة التقريبية؟

تمارين

(٥-٢١) إذا تم ترميز المتغير بحيث يصبح $\overline{X}=0$ وانطبق نموذج الانحدار بأخطاء طبيعية (3.1)، هل يكون b_0 و b_1 مستقلين؟ هل فترتا الثقة المشتركتان لـ β_0 و β_1 مستقلتان؟.

(٥-٢٢) استنبط تصميما لمتراجحة بونفرّوني ($5.3a$) إلى حالة ثلاث عبارات، كل منها بمعامل ثقة $(1 - \alpha)$.

(٥-٢٣) أثبت في حالة خط انحدار المربعات الدنيا عبر الأصل (5.16) أن $\sum X_i e_i = 0$.

(٥-٢٤) أثبت أن $\hat{Y}$ كما هي معرفة في (5.16) للانحدار الخطّي عبر الأصل مقدّر غير منحاز لـ $E\{Y\}$.

(٥-٢٥) استنبط الصيغة الخاصة بـ $S^2\{\hat{Y}_h\}$ المعطاة في جدول (٥-١) لنموذج انحدار خطّي عبر الأصل.

مشاريع

(٥-٢٦) عُد إلى مجموعة بيانات **SMSA** والمشروع (٢-٤٢) اعتبر علاقة الانحدار لعدد الأطباء الممارسين إلى عدد السكان الإجمالي.

ا - أوجد فترتي ثقة بونفرّوني المشتركة لـ β_0 و β_1 مستخدما 95 بالمائة معامل ثقة عائلة.

ب - يقترح باحث أن β_0 ينبغي أن تكون 400 - وأن β_1 ينبغي أن تكون 2.25.

هل تدعم فترتا الثقة المشتركة في الجزء (أ) وجهة النظر هذه؟ ناقش.

جـ ـ من المرغوب تقدير العدد المتوقع للأطباء العاملين من أجل **SMSA** بعدد سكان كلي $X = 500$، 1000 و5000 من الآلاف بمعامل ثقة عائلة 0.90 أي طريقة أكثر كفاءة هنا بونفرّوني أم ووركنج ـ هوتلّينج؟

د ـ أوجد عائلة التقديرات بفترة المطلوبة في الجزء (جـ) مستخدما الطريقة الأكثر كفاءة. فسّر فترات ثقتك.

(٥ـ٢٧) عد إلى مجموعة بيانات **SENIC** والمشروع (٢-٤٤) اعتبر علاقة الانحدار لمعدّل طول الإقامة إلى خطورة الإصابة.

ا ـ أوجد فترتي ثقة بونفرّوني المشتركة لـ β_0 و β_1 مستخدما 90 بالمائة معامل ثقة عائلة.

ب ـ اقترح باحث أن β_0 ينبغي أن تكون تقريبا 7 وينبغي أن تكون β_1 تقريبا 1. هل تدعم فترتا الثقة المشتركة في الجزء (ا) هذا التوقع؟ ناقش

جـ ـ من المرغوب تقدير إقامة المستشفى المتوقعة لأشخاص خطورة إصابتهم X=2,3,4,5 وذلك بمعامل ثقة عائلة 0.90 أي طريقة أكثر كفاءة هنا، ووركنج ـ هوتلّنج أم بونفيروني؟

د ـ أوجد عائلة التقديرات بفترة المطلوبة في الجزء (جـ) مستخدما الطريقة الأكثر كفاءة، فسّر فترات ثقتك.

طريقة المصفوفة لتحليل الانحدار الخطّي البسيط

يستخدم جبر المصفوفات على نطاق واسع في التحليل الرياضي والتحليل الإحصائي. وأسلوب المصفوفات ضرورة عملية في تحليل الانحدار المتعدّد، لأنه يسمح بالعرض الرمزي المختصر لنظم معادلات شاملة ولمصفوفات كثيرة من البيانات ولتطبيق عمليات على هذه الرموز بصورة فعّالة.

في هذا الفصل، نتابع أولا مقدمة مختصرة في جبر المصفوفات. (يمكن العثور على معالجة أتم لجبر المصفوفات في كتب متخصصة مثل مرجع [6.1]). ومن ثم نطبّق طرق المصفوفات على نموذج الانحدار الخطّي البسيط الذي ناقشناه في فصول سابقة. في حين لانحتاج حقيقة لجبر المصفوفات في نموذج الانحدار البسيط فإن تطبيق طرق المصفوفات على هذه الحالة يشكّل نقلة مفيدة إلى الانحدار المتعدّد، الذي سوف نتابعه في القسم الثاني من الجزء الأول.

قد يرغب القرّاء الملمّين بجبر المصفوفات إلقاء نظرة سريعة على الأجزاء التمهيدية لهذا الفصل ثم يركّزون على الأجزاء المتعلقة باستخدام طرق المصفوفات في تحليل الانحدار.

(٦-١) المصفوفات

تعريف مصفوفة

المصفوفة ترتيب مستطيل لعناصر منظمة في صفوف وأعمدة وكمثال لمصفوفة

	عمود 1	عمود 2
صف 1	16,000	23
صف 2	33,000	47
صف 3	21,000	35

عناصر هذه المصفوفة بـالذات هي أعداد تمثّل الدخل (العمود 1) و العمر (العمـود 2) لثلاثـة أشـخاص. والعنـاصر منظمـة وفقـا للصـف (شـخص) ووفقـا للعمـود (صفـة الشـخص). وهكـذا، يمثـل العنصـر في الصـف الأول والعمـود الأول (16,000) دخـل الشخص الأول. ويمثل العنصر في الصف الأول والعمـود الثـاني (23) عمـر الشـخص الأول، وبعد المصفوفة هو 2×3 يعني، 3 صفوف وعمودين. وإذا أردنـا تقديـم دخـل وعمر 1,000 شخص في مصفوفـة لهـا هيئـة كهيئـة المصفوفـة السـابقة، فسـنحتاج إلى مصفوفة $2 \times 1{,}000$.

وكأمثلة أخرى لمصفوفات نجد:

$$\begin{bmatrix} 1 & 0 \\ 5 & 10 \end{bmatrix} \quad \begin{bmatrix} 4 & 7 & 12 & 16 \\ 3 & 15 & 9 & 8 \end{bmatrix}$$

وأبعاد هاتين المصفوفتين هي 2×2 و2×4، على الترتيب. لاحظ أننا نحـدد دائمـا عـدد الصفوف أولا ثم عدد الأعمدة عند إعطاء بعد المصفوفة.

وكما هو في الجبر العادي، فقد نستخدم رموزا للتعبير عن عناصر مصفوفة:

$$\begin{array}{cc} & \begin{array}{ccc} j=1 & j=2 & j=3 \end{array} \\ \begin{array}{c} i=1 \\ i=2 \end{array} & \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \end{array}$$

لاحظ أن الدليل الأول يحدد رقم الصف ويحدد الثاني رقم العمود. وسنسـتخدم الرمـز العام a_{ij} للعنصر في الصف i والعمود j وفي مثالنا أعلاه $i = 1,2$ و $j = 1,2,3$.

ويمكن الإشارة للمصفوفـة برمـز مثـل **A**، **X** أو **Z**. والرمـز معطـى بخـط غـامق للتذكير بأنه يشير إلى مصفوفة، وهكذا، يمكننا تعريف المصفوفة أعلاه كما يلي:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

والإشارة إلى المصفوفة **A** يعني الإشارة إلى المصفوفة 2×3 المعطاة آنفا.

وكرمز آخر للمصفوفة **A** المعطاة آنفا نجد:

$$\mathbf{A} = [a_{ij}] \qquad i = 1,2; j = 1,2,3$$

وهو يتجنب الحاجة إلى كتابة جميع عناصر المصفوفة وذلك بعرض العنصر العام فقـط.

ويمكن استخدام هذا الرمز فقط، طبعا، عندما تكون عناصر المصفوفة رموزا.

والخلاصة، يمكن تمثيل مصفوفة بـ r صفا و c عمودا إما بعرضها كاملة:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \ldots a_{1j} \ldots a_{1c} \\ a_{21} & a_{22} \ldots a_{2j} \ldots a_{2c} \\ \cdot & \cdot \quad \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \quad \cdot \\ a_{i1} & a_{i2} \ldots a_{ij} \ldots a_{ic} \\ \cdot & \cdot \quad \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \quad \cdot \\ a_{r1} & a_{r2} \ldots a_{rj} \ldots a_{rc} \end{bmatrix} \tag{6.1}$$

أو بالشكل المختصر:

$$\mathbf{A} = [\alpha_{ij}] \quad i = 1,\ldots, r\,; j = 1,\ldots,c$$

أو ببساطة برمز مكتوب بخط غامق مثل **A**.

تعليقات

١ـ لا تفكر في المصفوفة وكأنها عدد. فهي مجموعة من العناصر مرتبة في مصفوفة. وفقط عندما يكون بُعد المصفوفة 1×1 فستتضمن المصفوفة عنصرا واحدا، وفي هذه الحالة يمكن التفكير فيها على أنها مصفوفة أو عدد.

٢ـ الشكل التالي ليس مصفوفة

$$\begin{bmatrix} & 14 & \\ & 8 & \\ 10 & & 15 \\ 9 & & 16 \end{bmatrix}$$

لأن الأعداد ليست مرتبة في أعمدة وصفوف

المصفوفة المربعة

يقال عن مصفوفة أنها مربعة إذا تساوى عدد صفوفها وعدد أعمدتها. وفيما يلي مثالان:

$$\begin{bmatrix} 4 & 7 \\ 3 & 9 \end{bmatrix} \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

المتجه

تسمى المصفوفة التي تحوي عمودا واحدا متجه عمود أو ببساطة متجها. وفيما يلي مثالان:

$$\mathbf{A}=\begin{bmatrix}4\\7\\10\end{bmatrix}\quad \mathbf{C}=\begin{bmatrix}c_1\\c_2\\c_3\\c_4\\c_5\end{bmatrix}$$

المتجه **A** هو مصفوفة 3×1 والمتجه **c** هو مصفوفة 5×1.

والمصفوفة التي تحوي صفا واحدا تسمى متجه صف. وكمثالين نذكر:

$$\mathbf{B}'=\begin{bmatrix}15 & 25 & 50\end{bmatrix}\qquad \mathbf{F}'=\begin{bmatrix}f_1 & f_2\end{bmatrix}$$

نستخدم رمز الفتحة (′) لمتجهات الصف لأسباب ستتضح قريبا. لاحظ أن متجه الصف **B′** مصفوفة 1×3 ومتجه الصف **F′** مصفوفة 1×2.

ويكفي دليل بمفرده للتمييز بين عناصر متجه.

منقول (مدوّر)

منقول مصفوفة **A** هو مصفوفة أخرى يُرمز لها بـ **A′**، والتي نحصل عليها بالمبادلة بين أعمدة المصفوفة **A** وصفوفها، كل عمود بالصف المقابل له.

وعلى سبيل المثال إذا كان:

$$\underset{3\times2}{\mathbf{A}}=\begin{bmatrix}2 & 5\\7 & 10\\3 & 4\end{bmatrix}$$

فإن المنقول هو **A′**:

$$\underset{2\times3}{\mathbf{A}'}=\begin{bmatrix}2 & 7 & 3\\5 & 10 & 4\end{bmatrix}$$

لاحظ أن العمود الأول من **A** هو الصف الأول من **A′** وبالمثل فإن العمود الثاني من **A** هو الصف الثاني من **A′** وفي المقابل، أصبح الصف الأول من **A** العمود الأول من **A′** وهكذا. لاحظ أن بُعد **A**، الموضح تحت الرمز أصبح معكوسا عند كتابة بُعد **A′**.

وكمثال آخر، اعتبر:

$$\underset{3\times1}{C}=\begin{bmatrix}4\\7\\10\end{bmatrix}\qquad \underset{1\times3}{C'}=\begin{bmatrix}4 & 7 & 10\end{bmatrix}$$

وهكذا فمنقول متجه عمود هو متجه صف والعكس بالعكس، وهـذا هـو السبب في استخدام الرمز $\mathbf{B}'$ سابقا لتمييز متجه الصف، إذ قد نعتبره منقول لمتجه عمود $\mathbf{B}$.

وبصورة عامة، لدينا:

$$\underset{r\times c}{\mathbf{A}}=\begin{bmatrix}a_{11}\cdots a_{1c}\\ \cdot \quad\quad \cdot\\ \cdot \quad\quad \cdot\\ \cdot \quad\quad \cdot\\ a_{r1}\cdots a_{rc}\end{bmatrix}=\begin{bmatrix}a_{ij}\end{bmatrix}\qquad i=1,\ldots,r;\ j=1,\ldots,c \tag{6.3}$$

تذيل صف   تذيل عمود

$$\underset{c\times r}{\mathbf{A}'}=\begin{bmatrix}a_{11}\cdots a_{r1}\\ \cdot \quad\quad \cdot\\ \cdot \quad\quad \cdot\\ \cdot \quad\quad \cdot\\ a_{1c}\cdots a_{rc}\end{bmatrix}=\begin{bmatrix}a_{ji}\end{bmatrix}\qquad j=1,\ldots,c;\ i=1,\ldots,r$$

تذيل صف   تذيل عمود

وهكذا نعثر على العنصر الواقع في الصف i والعمود j مـن $\mathbf{A}$ في الصـف J والعمـود i من $\mathbf{A}'$.

تساوي المصفوفات

يقال إن المصفوفتين $\mathbf{A}$ و $\mathbf{B}$ متساويتان إذا كـان لهمـا البعـد نفسـه وإذا تسـاوت العناصر المتقابلة. وعلى سبيل المثال:

$$\underset{3\times1}{\mathbf{A}}=\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}\qquad \underset{3\times1}{\mathbf{B}}=\begin{bmatrix}4\\7\\3\end{bmatrix}$$

فإن $\mathbf{B}=\mathbf{A}$ يتضمن:

$$a_1=4$$
$$a_2=7$$
$$a_3=3$$

وبالمثل، إذا كان:

$$\underset{3\times2}{\mathbf{A}}=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \qquad \underset{3\times2}{\mathbf{B}}=\begin{bmatrix} 17 & 2 \\ 14 & 5 \\ 13 & 9 \end{bmatrix}$$

فإن **A = B** يعني أن

$$\begin{array}{ll} a_{11}=17 & a_{12}=2 \\ a_{21}=14 & a_{22}=5 \\ a_{31}=13 & a_{32}=9 \end{array}$$

أمثلة انحدار

في تحليل الانحدار، المتجه **Y** هو مصفوفة أساسية، يتألف من الـ n مشاهدة في المتغير التابع:

$$\underset{n\times1}{\mathbf{Y}}=\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} \tag{6.4}$$

لاحظ أن المنقول **Y′** هو متجه الصف:

$$\underset{1\times n}{\mathbf{Y}'}=\begin{bmatrix} Y_1 & Y_2 & \dots Y_n \end{bmatrix} \tag{6.5}$$

والمصفوفة الأساسية الأخرى في تحليل الانحدار هي المصفوفة **X** والتي نعرّفها كما يلي في تحليل انحدار خطّي بسيط:

$$\underset{n\times2}{\mathbf{X}}=\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_n \end{bmatrix} \tag{6.6}$$

وتتألف المصفوفة **X** من عمود جميع عناصره 1 و عمود يحتوي على القيم الـ n للمتغير المستقل **X** لاحظ أن منقول **X** :

$$\underset{2\times n}{\mathbf{X}'} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \qquad (6.7)$$

وفي مثال حجم الدفعة لشركة وستوود، نجد المصفوفتين **Y** و**X** (الجـدول ٢-١) كما يلي:

$$\underset{10\times 1}{\mathbf{Y}} = \begin{bmatrix} 73 \\ 50 \\ . \\ . \\ . \\ 132 \end{bmatrix} \qquad \underset{10\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & 30 \\ 1 & 20 \\ . & . \\ . & . \\ . & . \\ 1 & 60 \end{bmatrix}$$

(٦-٢) جمع وطرح المصفوفات

يتطلب جمع وطرح مصفوفتين أن يكون لهما البُعْد نفسه. ومجموع أو فرق مصفوفتين هو مصفوفة أخرى كل عنصر من عناصرها يتألف من مجموع، أو فرق، العنصرين المتقابلين من المصفوفتين. افترض أن:

$$\underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \qquad \underset{3\times 2}{\mathbf{B}} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$$

فعندئذ:

$$\underset{3\times 2}{\mathbf{A}+\mathbf{B}} = \begin{bmatrix} 1+1 & 4+2 \\ 2+2 & 5+3 \\ 3+3 & 6+4 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 4 & 8 \\ 6 & 10 \end{bmatrix}$$

وبصورة مماثلة فإن:

$$\underset{3\times 2}{\mathbf{A}-\mathbf{B}} = \begin{bmatrix} 1-1 & 4-2 \\ 2-2 & 5-3 \\ 3-3 & 6-4 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 0 & 2 \end{bmatrix}$$

وبصورة عامة، إذا كان:

$$\underset{r\times c}{\mathbf{A}} = [a_{ij}] \qquad \underset{r\times c}{\mathbf{A}} = [b_{ij}] \qquad i = 1,\dots,r; j = 1,\dots,c$$

فعندئذ:

$$\underset{r\times c}{\mathbf{A}+\mathbf{B}}=\left[a_{ij}+b_{ij}\right] \quad \text{و} \quad \underset{r\times c}{\mathbf{A}-\mathbf{B}}=\left[a_{ij}-b_{ij}\right] \tag{6.8}$$

وتُعمَّم العلاقة (6.8) بطريقة واضحة إلى جمع أكثر من مصفوفتين. لاحظ كذلك أن $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$، كما في الجبر العادي.

مثال انحدار

يمكن كتابة نموذج الانحدار:

$$Y_i = E\{Y_i\} + \varepsilon_i \qquad i = 1,\ldots,n$$

بصورة متراصة باستخدام رمز المصفوفة. دعنا نعرف أولا متجه متوسطات الاستجابة:

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}}=\begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \cdot \\ \cdot \\ \cdot \\ E\{Y_n\} \end{bmatrix} \tag{6.9}$$

ومتجه حدود الخطأ:

$$\underset{n\times 1}{\boldsymbol{\varepsilon}}=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \tag{6.10}$$

متذكرين تعريف متجه المشاهدات $\mathbf{Y}$ في (6.4)، يمكن كتابة نموذج الانحدار كما يلي:

$$\underset{n\times 1}{\mathbf{Y}}=\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}}+\underset{n\times 1}{\boldsymbol{\varepsilon}}$$

لأن:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ .. \\ . \\ . \\ E\{Y_n\} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\}+\varepsilon_1 \\ E\{Y_2\}+\varepsilon_2 \\ .. \\ . \\ . \\ E\{Y_n\}+\varepsilon_n \end{bmatrix}$$

وهكذا يساوي متجه المشاهدات **Y** مجموع متجهين، متجه يحوي القيم المتوقعة وآخر يحوي حدود الخطأ.

(٦-٣) ضرب المصفوفات

ضرب المصفوفة بعدد سلّمي

العدد السلّمي هو عدد عادي أو رمز يمثل عددا. ولضرب مصفوفة بعدد سلّمي، يُضرب كل عنصر من المصفوفة بالعدد السلّمي. فمثلا، افترض أن المصفوفة معطاة بـ:

$$\mathbf{A} = \begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix}$$

فعندئذ **4A** حيث 4 عدد سلّمي، يساوي:

$$4\mathbf{A} = 4\begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 8 & 28 \\ 36 & 12 \end{bmatrix}$$

وبالمثل، $\lambda\mathbf{A}$ يساوي:

$$\lambda\mathbf{A} = \lambda\begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 2\lambda & 7\lambda \\ 9\lambda & 3\lambda \end{bmatrix}$$

حيث ترمز λ لعدد سلّمي.

عند وجود عامل مشترك لكل عنصر، يمكن أخذ هذا العامل خارج المصفوفة ومعاملته كعدد سلّمي. فمثلا

$$\begin{bmatrix} 9 & 27 \\ 15 & 18 \end{bmatrix} = 3\begin{bmatrix} 3 & 9 \\ 5 & 6 \end{bmatrix}$$

وبالمثل:

$$\begin{bmatrix} \frac{5}{\lambda} & \frac{2}{\lambda} \\ \frac{3}{\lambda} & \frac{8}{\lambda} \end{bmatrix} = \frac{1}{\lambda}\begin{bmatrix} 5 & 2 \\ 3 & 8 \end{bmatrix}$$

وعموما، إذا كان $\mathbf{A} = [a_{ij}]$ و λ عدد سلّمي فلدينا.

$$\lambda \mathbf{A} = \mathbf{A}\lambda = [\lambda a_{ij}] \tag{6.11}$$

ضرب مصفوفة بمصفوفة

قد يبدو ضرب مصفوفة بمصفوفة معقّدا بعض الشيء في البداية، ولكن قليلا من التمرين سيجعله عملية سهلة.

اعتبر المصفوفتين:

$$\underset{2\times2}{\mathbf{A}} = \begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} \qquad \underset{2\times2}{\mathbf{B}} = \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix}$$

فسيكون حاصل الضرب (جُداء) **AB** مصفوفة 2×2 نحصل على عناصرها بإيجاد الجداءات المتقاطعة لصفوف من **A** بأعمدة من **B** ثم جمع الجداءات المتصالبة. فمثلا، لإيجاد العنصر في الصف الأول والعمود الأول من حاصل الضرب **AB**، فيقتصر عملنا على الصف الأول من **A** والعمود الأول من **B**، وذلك كما يلي:

$$\begin{matrix} & \mathbf{A} & \mathbf{B} & & \mathbf{AB} \\ \text{صف 1} & \begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} & \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix} & \text{صف 1} & \begin{bmatrix} 33 & \\ & \end{bmatrix} \\ \text{صف 2} & & \text{عمود 1} \quad \text{عمود 2} & & \text{عمود 1} \end{matrix}$$

ونأخذ الجداءات المتصالبة ثم نجمع:

$$2(4) + 5(5) = 33$$

العدد 33 هو العنصر في الصف الأول والعمود الأول من المصفوفة **AB**.

ولإيجاد العنصر في الصف الأول والعمود الثاني من **AB**، نتناول الصف الأول من **A** والعمود الثاني من **B**:

$$\begin{matrix} & \mathbf{A} & \mathbf{B} & & \mathbf{AB} \\ \text{صف 1} & \begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} & \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix} & \text{صف 1} & \begin{bmatrix} 33 & 52 \\ & \end{bmatrix} \\ \text{صف 2} & & \text{عمود 1} \quad \text{عمود 2} & & \text{عمود 1} \end{matrix}$$

ومجموع الجداءات المتصالبة هو:

$$2(6) + 5(8) = 52$$

وباستمرار هذه العملية، نجد حاصل الضرب على الشكل:

$$\underset{2\times3}{\mathbf{AB}} = \begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix} = \begin{bmatrix} 33 & 52 \\ 21 & 32 \end{bmatrix}$$

دعنا نعتبر مثالا آخر:

$$\underset{2\times2}{\mathbf{A}} = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 5 & 8 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$$

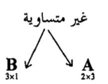
$$\underset{2\times2}{\mathbf{AB}} = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 5 & 8 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 26 \\ 41 \end{bmatrix}$$

عند الحصول على حاصل الضرب **AB**، نقول إن **A** مضروبة عن اليمين بـ **B** أو **B** مضروبة عن اليسار بـ **A**. والسبب في هذا المصطلح المحكم هو أن قواعد الضرب إلى الجبر الاعتيادي لاتنطبق في جبر المصفوفات. ففي الجبر المعتاد $xy = yx$، وفي جبر المصفوفات $\mathbf{AB} \neq \mathbf{BA}$ عادة. وفي الحقيقة حتى لو كان حاصل الضرب **AB** معرّفا، فقد لايكون حاصل الضرب **AB** معرّفا على الإطلاق.

وبصورة عامة، يكون حاصل الضرب **AB** معرفا فقط عندما يتساوى عدد الأعمدة في **A** مع عدد الصفوف في **B** بحيث تتوافر حدود متقابلة في الجداء المتصالب، وهكذا، ففي مثالينا السابقين: لدينا

متساويان	متساويان
$\underset{2\times3}{\mathbf{A}} \quad \underset{3\times1}{\mathbf{B}} = \underset{2\times1}{\mathbf{AB}}$	$\underset{2\times2}{\mathbf{A}} \quad \underset{2\times2}{\mathbf{B}} = \underset{2\times2}{\mathbf{AB}}$
بعد الضرب	بعد الضرب

لاحظ أن بُعد حاصل الضرب **AB** يعطى بعدد الصفوف في **A** وعدد الأعمدة في **B**. ولاحظ كذلك أنه في الحالة الثانية سوف لن يكون حاصل الضرب **AB** معرفا لأن عدد الأعمدة في **B** لايساوي عدد الصفوف في **A**.



وإليك مثال آخر لضرب المصفوفات.

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

وبصورة عامة، إذا كان بُعد **A** هو $r \times c$ وكان بُعد **B** هو $c \times s$ فإن بُعد حاصل الضرب **AB** هو $r \times s$ وعنصرها الواقع في الصف i والعمود j هو:

$$\sum_{k=1}^{c} a_{ik} b_{kj}$$

بذلك يكون:

$$\underset{r \times s}{\mathbf{AB}} = \left[\sum_{k=1}^{c} a_{ik} b_{kj} \right] \quad i = 1,\ldots,r \,;\, j = 1,\ldots, s \tag{6.12}$$

وهكذا، ففي المثال السابق يكون العنصر الواقع في الصف الأول والعمود الثاني من حاصل الضرب **AB** هو:

$$\sum_{k=1}^{3} a_{1k} b_{k2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}$$

وقد وجدناه، في الحقيقة، بأخذ الجداء المتصالب لعناصر الصف الأول من **A** بعناصر العمود الثاني من **B** ثم جمعها.

أمثلة إضافية

١- $\begin{bmatrix} 4 & 2 \\ 5 & 8 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 4a_1 + 2a_2 \\ 5a_1 + 8a_2 \end{bmatrix}$

٢- $\begin{bmatrix} 2 & 3 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 2^2 + 3^2 + 5^2 \end{bmatrix} = \begin{bmatrix} 38 \end{bmatrix}$

حاصل الضرب هنا هو مصفوفة 1 × 1، وهي تكافيء عددا سلميا. وهكذا فمصفوفة حاصل الضرب هنا تساوي العدد 38.

-٣
$$\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \beta_0 + \beta_1 X_3 \end{bmatrix}$$

أمثلة المحدار. حاصل الضرب الذي نحتاجه كثيرا هو **Y′Y**، حيث **Y** متجه المشاهدات في المتغير التابع كما عُرف في (6.4):

$$\underset{1\times1}{\mathbf{Y'Y}} = \begin{bmatrix} Y_1 \, Y_2 \, . \, . \, . \, Y_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} Y_1^2 + Y_2^2 + \ldots + Y_n^2 \end{bmatrix} = \begin{bmatrix} \sum Y_i^2 \end{bmatrix} \quad (6.13)$$

لاحظ أن **Y′Y** هو مصفوفة 1 × 1 أو عدد سلمي. وهكذا نجد طريقة متراصة لكتابة حدود مجموع المربعات: $\mathbf{Y'Y} = \sum Y_i^2$

وسوف نحتاج كذلك لـ **X′ X** وهو مصفوفة 2 × 2:

$$\underset{2\times2}{\mathbf{X'X}} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ X_1 & X_2 & \ldots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & \\ . & \\ . & \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \quad (6.14)$$

و**X′Y** وهو مصفوفة 1 × 2:

$$\underset{2\times1}{\mathbf{X'Y}} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ X_1 & X_2 & \ldots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \quad (6.15)$$

(٦-٤) أنواع خاصة من المصفوفات

تظهر أنواع خاصة بعينها من المصفوفات بصورة منتظمة في تحليل الانحدار، وسوف نعرض هنا أكثرها أهمية.

المصفوفة المتناظرة

إذا كان $\mathbf{A} = \mathbf{A}'$ يقال أن $\mathbf{A}$ متناظرة، وهكذا فالمصفوفة $\mathbf{A}$ التالية هي مصفوفة متناظرة:

$$\underset{3\times3}{\mathbf{A}} = \begin{bmatrix} 1 & 4 & 6 \\ 4 & 2 & 5 \\ 6 & 5 & 3 \end{bmatrix} \qquad \underset{3\times3}{\mathbf{A}'} = \begin{bmatrix} 1 & 4 & 6 \\ 4 & 2 & 5 \\ 6 & 5 & 3 \end{bmatrix}$$

ومن الواضح أن المصفوفة المتناظرة هي بالضرورة مصفوفة مربعة. تظهر المصفوفات المتناظرة تقليديا في تحليل الانحدار عندما نضرب عن اليسار مصفوفة $\mathbf{X}$ مثلا بمنقولها $\mathbf{X}'$. والمصفوفة الناتجة $\mathbf{X'X}$ ، متناظرة، كما يمكن رؤيته بسهولة من *(6.14)*.

المصفوفة القطرية

المصفوفة القطرية هي مصفوفة مربعة جميع عناصرها غير القطرية أصفار، مثل:

$$\underset{3\times3}{\mathbf{A}} = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix} \qquad \underset{4\times4}{\mathbf{B}} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

وسوف لا نكتب عادة جميع الأصفار في مصفوفة قطرية، فنقدمها على الشكل:

$$\underset{3\times3}{\mathbf{A}} = \begin{bmatrix} a_1 & & 0 \\ & a_2 & \\ 0 & & a_1 \end{bmatrix} \qquad \underset{4\times4}{\mathbf{B}} = \begin{bmatrix} 4 & & & 0 \\ & 1 & & \\ & & 10 & \\ 0 & & & 5 \end{bmatrix}$$

نوعان مهمان من المصفوفات القطرية هما مصفوفة الوحدة ومصفوفة عدد سلّمي.

مصفوفة الوحدة. مصفوفة الوحدة أو المصفوفة الواحدية ويرمز لها بـ $\mathbf{I}$ هي مصفوفة كل عنصر من عناصر قطرها الرئيس هو الواحد. وضرب أي مصفوفة مربعة $\underset{r\times r}{\mathbf{A}}$ عن اليسار أو عن اليمين بـ $r\times r$ مصفوفة وحدة $\mathbf{I}$ يترك $\mathbf{A}$ بدون تغيير. فمثلا:

$$\mathbf{IA}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

وبالمثل لدينا:

$$\mathbf{AI}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}$$

لاحظ أن مصفوفة الوحدة $\mathbf{I}$ تقابل إذن العدد 1 في الجبر العادي، حيث لدينا:

$1.\, x = x.1 = x$.

ولدينا لمصفوفة $\underset{r\times r}{\mathbf{A}}$:

$$\mathbf{AI}=\mathbf{IA}=\mathbf{A} \tag{6.16}$$

وهكذا يمكن إدخال أو حذف مصفوفة الوحدة من عبارة مصفوفية حينما يكون ذلك ملائما.

مصفوفة عدد سلّمي. مصفوفة عدد سلّمي هو مصفوفة قطرية جميع عناصر قطرها الرئيس متساوية. وكمثالين لمصفوفتي عدد سلّمي نذكر:

$$\begin{bmatrix}2&0\\0&2\end{bmatrix} \quad , \quad \begin{bmatrix}\lambda&0&0\\0&\lambda&0\\0&0&\lambda\end{bmatrix}$$

ويمكن التعبير عن مصفوفة عدد سلّمي على الشكل $\lambda\mathbf{I}$ حيث λ عدد سلّمي فمثلا،

$$\begin{bmatrix}2&0\\0&2\end{bmatrix}=2\begin{bmatrix}1&0\\0&1\end{bmatrix}=2\mathbf{I}$$

$$\begin{bmatrix}\lambda&0&0\\0&\lambda&0\\0&0&\lambda\end{bmatrix}=\lambda\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}=\lambda\mathbf{I}$$

وضرب $r \times r$ مصفوفة **A** بـ $r \times r$ مصفوفة عدد سلّمي $\lambda\mathbf{I}$ يكافىء ضرب **A** بالعدد السلّمي λ.

متجه ومصفوفة جميع عناصرهما الواحد

سنرمز لمتجه عمود جميع عناصره 1 بـ **1**

$$\underset{r\times 1}{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \tag{6.17}$$

وسيُرمز لمصفوفة مربعة جميع عناصرها 1 بـ **J** :

$$\underset{r\times r}{\mathbf{J}} = \begin{bmatrix} 1 & \cdots & 1 \\ 1 & & 1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 1 & \cdots & 1 \end{bmatrix} \tag{6.18}$$

فمثلا، لدينا:

$$\underset{3\times 1}{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \underset{3\times 3}{\mathbf{J}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

لاحظ أننا نحصل، من أجل $n \times 1$ متجه **1** على:

$$\underset{1\times 1}{\mathbf{1}'\mathbf{1}} = [1 \;.\;.\;.\; 1] \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} = [n] = n$$

و:

$$\underset{n\times n}{\mathbf{11'}}=\begin{bmatrix}1\\ .\\ .\\ .\\ 1\end{bmatrix}[1\ .\ .\ .1]=\begin{bmatrix}1 & ...1\\ . & .\\ . & .\\ . & .\\ 1 & ...1\end{bmatrix}=\underset{n\times n}{\mathbf{J}}$$

متجه صفري

المتجه الصفري: متجه يحوي أصفارا فقط. وسيرمز لمتجه عمود صفري بـ ($\mathbf{0}$):

$$\underset{r\times 1}{\mathbf{0}}=\begin{bmatrix}0\\ 0\\ .\\ .\\ .\\ 0\end{bmatrix} \tag{6.19}$$

على سبيل المثال، لدينا:

$$\underset{3\times 1}{\mathbf{0}}=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}$$

(٦-٥) الاعتماد الخطّي ورتبة مصفوفة

اعتماد خطّي

اعتبر المصفوفة التالية:

$$\mathbf{A}=\begin{bmatrix}1 & 2 & 5 & 1\\ 2 & 2 & 10 & 6\\ 3 & 4 & 15 & 1\end{bmatrix}$$

ولنفكر الآن في أعمدة هذه المصفوفة كمتجهات. وهكذا، ننظر إلى $\mathbf{A}$ كما لو كانت مكونة من أربعة متجهات عمود، وقد اتّفق هنا أن كانت الأعمدة على صلة خاصة بعضها ببعض. إذ نلاحظ أن متجه العمود الثالث من مضاعفات متجه العمود الأول:

$$\begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix} = 5\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

نقول إن أعمدة **A** مرتبطة خطّيا، فهي تنطوي على نافلة من المعلومــات، كمـا يُقـال، حيث يمكن الحصول على أحد الأعمدة كتركيب خطّي في الأعمدة الأخرى.

ونعرّف مجموعة من متجهات العمود على أنها مرتبطة خطّيا إذا أمكن التعبير عن أحد المتجهات كتركيب خطّي في الأعمـدة الأخـرى. وإذا لم نسـتطع التعبـير عـن أي متجه في المجموعة كتركيب خطّي في المتجهات الأخرى، فنعـرّف مجموعـة المتجهـات عندئذ بأنها مستقلة خطّيا. وفيما يلي تعريف أكثر عمومية، ولكنه مكافىء لاسـتقلال c متجه عمود $\mathbf{C}_1$،..،$\mathbf{C}_c$ في مصفوفة $r \times c$ عندما يمكـن إيجـاد **c** عـددا سـلّميا λ_1،...،λ_c ليست جميعها أصفارا بحيث يكون:

$$\lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2 + \ldots + \lambda_c\mathbf{C}_c = \mathbf{0} \qquad (6.20)$$

حيث يرمز **0** إلى متجه عمود صفري، فتكون متجهات العمود، وعددها c، مرتبطة خطيـا. وإذا كـانت المجموعـة الوحيـدة مـن الأعـداد السـلّمية الـتي تحقــق المســاواة هــي المجموعــة $\lambda_1=0,..,\lambda_c = 0$ فإن مجموعة متجهات العمود، وعددها c، تكون مستقلة خطيا.

وللتوضيح نجد في مثالنا أن $\lambda_1 = 5$، $\lambda_2 = 0$، $\lambda_3 = -1$ و $\lambda_4 = 0$ تؤدي إلى:

$$5\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 0\begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} - 1\begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 6 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

وبالتالي تكون متجهات العمود مرتبطة خطيا ـ لاحظ هنا أن بعضا من الــ λ_j يســاوي الصفر. ويتطلب الارتباط الخطّي فقط أن λ_j ليست جميعها أصفارا.

رتبة مصفوفة

تُعرَّف رتبة مصفوفة بأنها أكبر عدد مــن الأعمـدة المسـتقلة في المصفوفـة، ونحـن نعلم أن رتبـة **A** في مثالنـا السـابق لايمكـن أن تكـون 4 لأن الأعمـدة الأربعـة مرتبطـة خطّيا، وعلى أي حال، نستطيع ايجـاد ٣ أعمـدة (1، 2 و4) مسـتقلة خطيـا. فلايمكـن

إيجاد أعداد سلّمية λ_1، λ_2 و λ_4 بحيث يكون $\lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2 + \lambda_4\mathbf{C}_4 = \mathbf{0}$ إلا إذا كان $\lambda_1=\lambda_2=\lambda_4=0$ وهكذا تكون رتبة **A** في مثالنا هي 3.

رتبة المصفوفة وحيدة، ويمكن تعريفها بصورة مكافئة كأكبر عدد من الصفوف المستقلة خطّيا. وينتج أن رتبة $r \times c$ مصفوفة لا يمكن أن تزيد على $min\ (r,c)$ (أصغر العددين r و c).

(٦-٦) معكوس مصفوفة

معكوس عدد، في الجبر العادي، هو مقلوبه، وهكذا، فمعكوس 6 هو $\frac{1}{6}$ وحاصل ضرب عدد بمعكوسه يساوي دائما 1:

$$6.\frac{1}{6}=1$$

$$x.\frac{1}{x}=x.x^{-1}=x^{-1}.x=1$$

في جبر المصفوفات، نجد أن معكوس مصفوفة **A** هو مصفوفة أخرى، نرمز لها بـ $\mathbf{A}^{-1}$، بحيث يكون:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \tag{6.21}$$

حيث **I** مصفوفة الوحدة. وهكذا تلعب المصفوفة **I**، من جديد، الدور نفسه الذي يلعبه الواحد في الجبر العادي. ومعكوس مصفوفة معرف فقط من أجل المصفوفات المربعة. وحتى في هذه المصفوفات قد لايكون لكثير من المصفوفات المربعة معكوس وإذا كان لمصفوفة معكوس بالفعل فإن هذا المعكوس وحيد.

أمثلة

١- معكوس المصفوفة:

$$\underset{2\times2}{\mathbf{A}}=\begin{bmatrix}2 & 4\\3 & 1\end{bmatrix}$$

هو:

$$\underset{2\times2}{\mathbf{A}}^{-1}=\begin{bmatrix}-.1 & .4\\.3 & -.2\end{bmatrix}$$

ذلك لأن:

$$\mathbf{A}^{-1}\mathbf{A}=\begin{bmatrix}-.1 & .4\\ .3 & -.2\end{bmatrix}\begin{bmatrix}2 & 4\\ 3 & 1\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

أو:

$$\mathbf{A}\mathbf{A}^{-1}=\begin{bmatrix}2 & 4\\ 3 & 1\end{bmatrix}\begin{bmatrix}-.1 & .4\\ .3 & -.2\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

٢- معكوس المصفوفة:

$$\underset{3\times 3}{\mathbf{A}}=\begin{bmatrix}3 & 0 & 0\\ 0 & 4 & 0\\ 0 & 0 & 2\end{bmatrix}$$

هو:

$$\underset{3\times 3}{\mathbf{A}}^{-1}=\begin{bmatrix}\frac{1}{3} & 0 & 0\\ 0 & \frac{1}{4} & 0\\ 0 & 0 & \frac{1}{2}\end{bmatrix}$$

ذلك لأن:

$$\mathbf{A}^{-1}\mathbf{A}=\begin{bmatrix}\frac{1}{3} & 0 & 0\\ 0 & \frac{1}{4} & 0\\ 0 & 0 & \frac{1}{2}\end{bmatrix}\begin{bmatrix}3 & 0 & 0\\ 0 & 4 & 0\\ 0 & 0 & 2\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

لاحظ أن معكوس مصفوفة قطرية هو مصفوفة قطرية تتألف ببساطة من مقلوبات العناصر القطرية.

ايجاد المعكوس

حتى هذه النقطة، أُعطي معكوس مصفوفة **A**، وقد جربنا فقط للتحقق من أنه المعكوس، برؤية ما إذا $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ أم لا. ولكن كيف نجد المعكوس، ومتى يوجد معكوس مصفوفة مربعة $r \times r$ إذا كانت رتبة المصفوفة r. ويُقال عن مثل هذه المصفوفة إنها **غير شاذة**. ويقال عن مصفوفة $r \times r$ رتبتها أقل من r **أنها شاذة**، وليس لها معكوس.

وقد يتطلب ايجاد معكوس مصفوفة كمية ضخمة من الحسابات. وسوف نتخذ أسلوبا في هذا الكتاب يقضي بإمكانية حساب معكوس مصفوفة 2 × 2 ومصفوفة 3×3 يدويا. ولأي مصفوفة كبيرة، يستخدم عادة الحاسب الآلي أو حاسبة يدوية مبرمجة لايجاد المعكوس، إلا إذا كانت المصفوفة من شكل معين مثل المصفوفة القطرية.
ويمكن تبيان أن معكوسي مصفوفتين 2 × 2 و 3×3 هما كالتالي:

١ – إذا كانت:

$$\underset{2\times 2}{\mathbf{A}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

فعندئذ:

$$\underset{2\times 2}{\mathbf{A}}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{d}{D} & \frac{-b}{D} \\ \frac{-c}{D} & \frac{a}{D} \end{bmatrix} \qquad (6.21)$$

حيث:

$$D = ad - bc$$

وتسمى D **محددة المصفوفة A**، ولو كانت **A** شاذة، فمحددتها ستساوي صفرا، ولا يوجد لها معكوس.

٢- إذا كانت:

$$\underset{3\times 3}{\mathbf{B}} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$$

فعندئذ :

$$\underset{3\times 3}{\mathbf{B}}^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}^{-1} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix} \qquad (6.23)$$

حيث :

$$A = (ek - fh) / Z \qquad B = -(bk - ch) / Z \qquad C = (bf - ce) / Z$$
$$D = -(dk - fg) / Z \qquad E = (ak - cg) / Z \qquad F = -(af - cd) / Z$$
$$G = (dh - eg) / Z \qquad H = -(ah - bg) / Z \qquad K = (ae - bd) / Z$$

و:

$$Z = a(ek - fh) - b(dk - fg) + c(dh - eg)$$

تسمى Z محددة المصفوفة **B**.

لنستخدم (6.22) لإيجاد المعكوس لـ:

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$$

فلدينا:

$$b = 4 \qquad a = 2$$
$$d = 1 \qquad c = 3$$
$$D = ad - bc = 2(1) - 4(3) = -10$$

وبالتالي:

$$\mathbf{A}^{-1} = \begin{bmatrix} \frac{1}{-10} & \frac{-4}{-10} \\ \frac{-3}{-10} & \frac{2}{-10} \end{bmatrix} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix}$$

كما وجدناه في مثال سابق.

وعندما نحصل على معكوس $\mathbf{A}^{-1}$ إما باستخدام الحسابات اليدوية أو من تشغيلة حاسب آلي، فمن الحكمة، عادة، حساب $\mathbf{AA}^{-1}$ للتحقق مما إذا كان حاصل الضرب يساوي مصفوفة الوحدة أم لا، سامحين بانحرافات طفيفة عن 0 و 1 كنتيجة لتدوير الأرقام العشرية.

مثال انحدار

مصفوفة المعكوس الأساسية التي نواجهها في تحليل الانحدار هو معكوس المصفوف **X'X** في (6.14):

$$\underset{2\times2}{\mathbf{X'X}} = \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}$$

وباستخدام القاعدة (6.22) لدينا:

$$a = n \qquad b = \Sigma X_i$$
$$c = \Sigma X_i \qquad d = \Sigma X_i^2$$

بحيث يكون:

$$D=n\sum X_i^2-(\sum X_i)(\sum X_i)=n\left[\sum X_i^2-\frac{(\sum X_i)^2}{n}\right]=n\sum(X_i-\overline{X})^2$$

وبالتالي:

$$\underset{2\times 2}{(\mathbf{X'X})^{-1}}=\begin{bmatrix}\dfrac{\sum X_i^2}{n\sum(X_i-\overline{X})^2} & \dfrac{-\sum X_i}{n\sum(X_i-\overline{X})^2}\\ \dfrac{-\sum X_i}{n\sum(X_i-\overline{X})^2} & \dfrac{n}{n\sum(X_i-\overline{X})^2}\end{bmatrix} \qquad (6.24)$$

وبما أن $\sum X_i=n\overline{X}$ ، فنستطيع تبسيط (6.24):

$$\underset{2\times 2}{(\mathbf{X'X})^{-1}}=\begin{bmatrix}\dfrac{\sum X_i^2}{n\sum(X_i-\overline{X})^2} & \dfrac{-\overline{X}}{\sum(X_i-\overline{X})^2}\\ \dfrac{-\overline{X}}{\sum(X_i-\overline{X})^2} & \dfrac{1}{\sum(X_i-\overline{X})^2}\end{bmatrix} \qquad (6.25)$$

استخدامات معكوس مصفوفة

في الجبر العادي، نحل معادلة من النوع:

$$5y = 20$$

بضرب طرفي المعادلة بمعكوس 5 ونعني:

$$\frac{1}{5}(5y)=\frac{1}{5}(20)$$

ونحصل بذلك على:

$$y=\frac{1}{5}(20)=4$$

وفي المقابل إذا كان لدينا المعادلة التالية في جبر المصفوفات:

$$\mathbf{AY = C}$$

فنضرب الطرفين من اليسار بـ $\mathbf{A}^{-1}$ مفترضين أن للمصفوفة **A** معكوسا:

$$\mathbf{A^{-1}\,AY= A^{-1}C}$$

وبما أن $\mathbf{A^{-1}AY = IY = Y}$ نجد:

$$\mathbf{Y = A^{-1}C}$$

لتوضيح هذا الاستخدام، افترض أن لدينا معادلتين متزامنتين:

$$2y_1 + 4y_2 = 20$$
$$3y_1 + y_2 = 10$$

فيمكن كتابتهما كالتالي برموز المصفوفات:

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

وحل هاتين المعادلتين عندئذ هو:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

وسبق أن وجدنا المعكوس المطلوب، ولذلك نجد:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix} \begin{bmatrix} 20 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

وبالتالي $y_1 = 2$ و $y_2 = 4$ تحققان هاتين المعادلتين.

(٦–٧) بعض النظريات الأساسية للمصفوفات

نضع هنا قائمة، وبدون اثبات، ببعض من النظريات الأساسية للمصفوفات والتي سنستفيدُ منها في عمل قادم.

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{6.26}$$

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \tag{6.27}$$

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \tag{6.28}$$

$$\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB} \tag{6.29}$$

$$\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B} \tag{6.30}$$

$$(\mathbf{A}')' = \mathbf{A} \tag{6.31}$$

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}' \tag{6.32}$$

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \tag{6.33}$$

$$(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}' \tag{6.34}$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{6.35}$$

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \tag{6.36}$$

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A} \tag{6.37}$$

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})' \tag{6.38}$$

(٦–٨) متجهات ومصفوفات عشوائية

المتجه العشوائي أو المصفوفة العشوائية تتضمن عناصر هي متغيرات عشوائية. وهكذا يكون متجه المشاهدات **Y** في (6.4) متجها عشوائيا لأن العناصر Y_i متغيرات عشوائية.

توقع متجه أو مصفوفة عشوائية

افترض أن لدينا $n = 3$ مشاهدات ونهتم بمتجه المشاهدات:

$$\underset{3\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

فالقيمة المتوقعة للمتجه $\mathbf{Y}$، ويرمز لها بـ $\mathbf{E}\{Y\}$، معرفة كالتالي:

$$\underset{3\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ E\{Y_3\} \end{bmatrix}$$

وهكذا، فإن القيمة المتوقعة لمتجه عشوائي هي متجه عناصره القيم المتوقعة للمتغيرات العشوائية التي تشكل عناصر المتجه العشوائي. وبالمثل، فإن توقع مصفوفة عشوائية هو مصفوفة عناصرها القيم المتوقعة للمتغيرات العشوائية المقابلة في المصفوفة الأصلية. وقد واجهنا متجه قيم متوقعة سابقا في (6.4).

وبصورة عامة، فإن توقع متجه عشوائي $\mathbf{Y}$ هو:

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \left[E\{Y_i\}\right] \qquad i = 1, \ldots, n \qquad (6.39)$$

ولمصفوفة عشوائية $\boldsymbol{Y}$ ببعد $n \times p$ يكون التوقع:

$$\underset{n\times p}{\mathbf{E}\{\mathbf{Y}\}} = \left[E\{Y_{ij}\}\right] \qquad i = 1, \ldots, n; j = 1, \ldots, p \qquad (6.40)$$

مثال انحدار. افترض أن عدد المشاهدات في تطبيق انحدار هو $n = 3$، فلكل من حدود الخطأ الثلاثة ε_1، ε_2 وε_3 توقع يساوي الصفر. ولمتجه الخطأ:

$$\underset{3\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

لدينا:

$$\underset{3\times 1}{\mathbf{E}\{\boldsymbol{\varepsilon}\}} = \underset{3\times 1}{\mathbf{0}}$$

لأن:

$$\mathbf{E}\{\varepsilon\} = \begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ E\{\varepsilon_3\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}$$

مصفوفة تباين ـ تغاير متجه عشوائي

اعتبر ثانية المتجه العشوائي $\mathbf{Y}$ المؤلف من المشاهدات الثلاث Y_1، Y_2 وY_3، فلكل متغير عشوائي تباين، $\sigma^2\{Y_i\}$، ولكل متغيرين عشوائيين تغاير $\sigma\{Y_i, Y_j\}$، ويمكن تجميع هذه في مصفوفة تسمى مصفوفة تباين ـ تغاير ويرمز لها بـ$\sigma^2\{\mathbf{Y}\}$:

$$\sigma^2\{\mathbf{Y}\} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \sigma\{Y_1, Y_3\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \sigma\{Y_2, Y_3\} \\ \sigma\{Y_3, Y_1\} & \sigma\{Y_3, Y_2\} & \sigma^2\{Y_3\} \end{bmatrix} \quad (6.41)$$

لاحظ أن التباينات واقعة على القطر الرئيس بينما نجد التغاير $\sigma\{\mathbf{Y}_i, \mathbf{Y}_j\}$ في الصف i والعمود j من المصفوفة. وهكذا يتواجد $\sigma\{Y_2, Y_1\}$ في الصف الثاني والعمود الأول ويتواجد في الصف الأول والعمود الثاني وتذكر طبعا، أن $\sigma\{Y_1, Y_2\}=\sigma\{Y_2, Y_1\}$ وبما أن $\sigma\{Y_i, Y_j\} = \sigma\{Y_j, Y_i\}$، بصورة عامة، من أجل $i \neq j$ فالمصفوفة $\sigma^2\{\mathbf{Y}\}$ متناظرة.

ويتبع بسهولة أن:

$$\sigma^2\{Y\} = E\left\{\left[Y - E\{Y\}\right]\left[Y - E\{Y\}\right]'\right\} \quad (6.42)$$

ولتوضيحنا هنا، لدينا:

$$\sigma^2\{\mathbf{Y}\} = \mathbf{E}\left\{\begin{bmatrix} Y_1 - E\{Y_1\} \\ Y_2 - E\{Y_2\} \\ Y_3 - E\{Y_3\} \end{bmatrix}\begin{bmatrix} Y_1 - E\{Y_1\} & Y_2 - E\{Y_2\} & Y_3 - E\{Y_3\} \end{bmatrix}\right\}$$

وبضرب المصفوفتين. ثم أخذ التوقعات، نجد:

الموقع في حاصل الضرب	الحد	القيمة المتوقعة
صف ١ عمود ١	$(Y_1 - E\{Y_1\})^2$	$\sigma^2\{Y_1\}$
صف ١ عمود ٢	$(Y_1 - E\{Y_1\})(Y_2 - E\{Y_2\})$	$\sigma\{Y_1, Y_2\}$
صف ١ عمود ٣	$(Y_1 - E\{Y_1\})(Y_3 - E\{Y_3\})$	$\sigma\{Y_1, Y_3\}$
صف ٢ عمود ١	$(Y_2 - E\{Y_2\})(Y_1 - E\{Y_1\})$	$\sigma\{Y_2, Y_1\}$
	etc.	etc.

ويقود هذا، طبعا، إلى مصفوفة تباين ـ تغاير في (6.41). تذكّر، عند أخذ التوقعات، تعاريف التباين والتغاير في (1.15) و(1.21) على الترتيب.

ولتعميم مصفوفة التباين ـ التغاير إلى $n \times 1$ متجه عشوائي **Y** نكتب:

$$\underset{n\times n}{\sigma^2\{\mathbf{Y}\}} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1,Y_2\} & \dots & \sigma\{Y_1,Y_n\} \\ \sigma\{Y_2,Y_1\} & \sigma^2\{Y_2\} & \dots & \sigma\{Y_2,Y_n\} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ ..\sigma\{Y_n,Y_1\} & \sigma\{Y_n,Y_2\} & \dots & \sigma^2\{Y_n\} \end{bmatrix} \quad (6.43)$$

لاحظ مرة أخرى أن $\sigma^2\{\mathbf{Y}\}$ مصفوفة متناظرة.

مثال انحدار. لنعد إلى المثال المستند إلى $n = 3$ مشاهدات. افترض أن لحدود الخطأ الثلاثة تباين ثابت $\sigma^2 = \{\varepsilon_i\}\sigma^2$، وأنها غير مرتبطة، أي أن $\sigma^2\{\varepsilon_i, \varepsilon_j\} = 0$ لـ $i \neq j$، فيمكننا عندئذ كتابة مصفوفة التباين ـ التغاير للمتجه العشوائي ε للمثال السابق كما يلي:

$$\underset{3\times 3}{\sigma^2\{\varepsilon\}} = \underset{3\times 3}{\sigma^2\mathbf{I}}$$

ذلك لأن:

$$\sigma^2\mathbf{I} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

لاحظ أن كل التباينات σ^2 وأن كل التغايرات صفر.

بعض النظريات الأساسية

كثيرا ما نواجه متجها عشوائيا **W** نحصل عليه بضرب المتجه العشوائي **Y** من اليسار بمصفوفة ثابتة **A** (مصفوفة عناصرها مثبتة)

$$\mathbf{W} = \mathbf{AY} \quad (6.44)$$

وفيما يلي بعض من النظريات الأساسية لهذه الحالة

$$\mathbf{E}\{\mathbf{A}\} = \mathbf{A} \quad (6.45)$$

$$\mathbf{E}\{\mathbf{W}\} = \mathbf{E}\{\mathbf{AY}\} = \mathbf{AE}\{\mathbf{Y}\} \quad (6.46)$$

$$\sigma^2\{\mathbf{W}\} = \sigma^2\{\mathbf{AY}\} = \mathbf{A}\sigma^2\{\mathbf{Y}\}\mathbf{A}' \quad (6.47)$$

حيث $\sigma^2\{\mathbf{Y}\}$ مصفوفة تباين ـ تغاير $\mathbf{Y}$.

مثال. كتوضيح بسيط لاستخدام هذه النظريات، اعتبر:

$$\underset{2\times1}{\mathbf{W}} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \underset{2\times2}{\mathbf{A}}\underset{2\times1}{\mathbf{Y}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} Y_1 - Y_2 \\ Y_1 + Y_2 \end{bmatrix}$$

فلدينا من (6.46):

$$\underset{2\times1}{\mathbf{E}\{\mathbf{W}\}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \end{bmatrix} = \begin{bmatrix} E\{Y_1\} - E\{Y_2\} \\ E\{Y_1\} + E\{Y_2\} \end{bmatrix}$$

ومن (6.47):

$$\underset{2\times2}{\sigma^2\{\mathbf{W}\}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2\{Y_1\} + \sigma^2\{Y_2\} - 2\sigma\{Y_1, Y_2\} & \sigma^2\{Y_1\} - \sigma^2\{Y_2\} \\ \sigma^2\{Y_1\} - \sigma^2\{Y_2\} & \sigma^2\{Y_1\} + \sigma^2\{Y_2\} + 2\sigma\{Y_1, Y_2\} \end{bmatrix}$$

وهكذا نجد:

$$\sigma^2\{W_1\} = \sigma^2\{Y_1 - Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} - 2\sigma\{Y_1, Y_2\}$$
$$\sigma^2\{W_2\} = \sigma^2\{Y_1 + Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} + 2\sigma\{Y_1, Y_2\}$$
$$\sigma\{W_1, W_2\} = \sigma\{Y_1 - Y_2, Y_1 + Y_2\} = \sigma^2\{Y_1\} - \sigma^2\{Y_2\}$$

(٦-٩) انحدار خطّي بسيط بدلالة المصفوفات

نحن الآن على استعداد لتطوير انحدار خطّي بسيط بدلالة المصفوفات. تذكر ثانية أننالن نقدّم أية نتائج جديدة، ولكننا سنعرض فقط النتائج التي حصلنا عليها سابقا بدلالة المصفوفات. وسنبدأ بنموذج الانحدار (3.1):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1, \ldots, n \qquad (6.48)$$

وهذا يتضمن:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned} \qquad (6.48a)$$

عرّفنا سابقا متجه المشاهدات **Y** في (6.4) والمصفوفة **X** في (6.6) والمتجه ε في (6.10). دعنا نعيد التعريفات ونعرّف المتجه β لمعاملات الانحدار:

$$\underset{n\times1}{\mathbf{Y}}=\begin{bmatrix}Y_1\\Y_2\\.\\.\\.\\Y_n\end{bmatrix}\quad \underset{n\times2}{\mathbf{X}}=\begin{bmatrix}1 & X_1\\1 & X_2\\. & .\\ & .\\ & .\\1 & X_n\end{bmatrix}\quad \underset{2\times1}{\boldsymbol{\beta}}=\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}\quad \underset{n\times1}{\boldsymbol{\varepsilon}}=\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\.\\.\\.\\\varepsilon_n\end{bmatrix} \tag{6.49}$$

يمكننا الآن كتابة (6.48a)، بصورة متراصّة، بدلالة المصفوفات وذلك كما يلي:

$$\underset{n\times1}{\mathbf{Y}}=\underset{n\times2}{\mathbf{X}}\;\underset{2\times1}{\boldsymbol{\beta}}+\underset{n\times1}{\boldsymbol{\varepsilon}} \tag{6.50}$$

لأن:

$$\begin{bmatrix}Y_1\\Y_2\\.\\.\\.\\Y_n\end{bmatrix}=\begin{bmatrix}1 & X_1\\1 & X_2\\. & .\\ & .\\ & .\\1 & X_n\end{bmatrix}\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\.\\.\\.\\\varepsilon_n\end{bmatrix}$$

$$=\begin{bmatrix}\beta_0+\beta_1X_1\\\beta_0+\beta_1X_2\\. \quad .\\ .\\ .\\\beta_0+\beta_1X_n\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\.\\.\\.\\\varepsilon_n\end{bmatrix}=\begin{bmatrix}\beta_0+\beta_1X_1+\varepsilon_1\\\beta_0+\beta_1X_2+\varepsilon_2\\. \quad .\\ .\\ .\\\beta_0+\beta_1X_n+\varepsilon_n\end{bmatrix}$$

ونلاحظ أن $\mathbf{X}\beta$ هو متجه القيم المتوقعة للمشاهدات Y_i لأن $E\{Y_i\}=\beta_0+\beta_1X_i$ وبالتالي:

$$\underset{n\times1}{\mathbf{E\{Y\}}}=\underset{n\times1}{\mathbf{X}\,\boldsymbol{\beta}} \tag{6.51}$$

حيث **E{Y}** معرّف في (6.9).

ويمكن النظر إلى العمود 1 في المصفوفة **X** كعمود مؤلف من المتغير الدمية $X_0\equiv1$ في نموذج الانحدار البديل (2.5):

$$Y_i=\beta_0X_0+\beta_1X_i+\varepsilon_i \qquad ; \qquad X_0\equiv1$$

و هكذا يمكن اعتبار المصفوفة **X** مصفوفة مؤلفة مـن متجـه عمـود للمتغـير الدميـة X_0 ومتجه عمود آخر يتألف من مشاهدات المتغير المستقل $\mathbf{X}_i$.

وبالنسبة لحدود الخطأ، يفترض نموذج الانحدار (3.1) أن $E\{\varepsilon_i\} = 0$ و $\sigma^2\{\varepsilon_i\} = \sigma^2$ وأن ε_i متغيرات عشوائية طبيعية مستقلة. ويصبح الشرط $E\{\varepsilon_i\}=0$ بدلالة المصفوفات:

$$\underset{n\times 1}{\mathbf{E}\{\varepsilon\}} = \underset{n\times 1}{\mathbf{0}} \tag{6.52}$$

لأن (6.52) تفيد أن:

$$\mathbf{E}\{\varepsilon\} = \begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ \cdot \\ \cdot \\ \cdot \\ E\{\varepsilon_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

والشرط أن لحدود الخطأ تباينا ثابتا σ^2 وأن جميع التغايرات $\sigma\{\varepsilon_i, \varepsilon_j\}$ صفـر مـن أجل $j \neq i$ (طالما أن الـ ε_i مستقلة)، هذا الشرط يُعبَّر عنه بدلالة المصفوفات من خـلال مصفوفة التباين ـ التغاير:

$$\underset{n\times n}{\sigma^2\{\varepsilon\}} = \underset{n\times n}{\sigma^2\mathbf{I}} \tag{6.53}$$

ذلك لأن (6.53) تفيد ما يلي:

$$\sigma^2\{\varepsilon\} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ \cdot & \cdot & \cdots & & \cdot \\ \cdot & \cdot & \cdots & & \cdot \\ \cdot & \cdot & \cdots & & \cdot \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

وهكذا، يكون نموذج انحدار الخطأ الطبيعي (3.1) بدلالة المصفوفات:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{6.54}$$

حيث:

ε متجه متغيرات عشوائية طبيعية مستقلة مع $\mathbf{E}\{\varepsilon\} = \mathbf{0}$ و $\sigma^2\{\varepsilon\} = \sigma^2\mathbf{I}$

(٦-١٠) تقدير المربعات الدنيا لمعالم الانحدار

المعادلات الناظمية

المعادلات الناظمية (2.9):

$$
\begin{aligned}
nb_0 + b_1 \sum X_i &= \sum Y_i \\
\mathbf{b}_0 \sum \mathbf{X_i} + \mathbf{b}_1 \sum \mathbf{X_i^2} &= \sum \mathbf{X_i Y_i}
\end{aligned} \tag{6.55}
$$

وبدلالة المصفوفات نجد:

$$\underset{2\times2}{\mathbf{X'X}}\ \underset{2\times1}{\mathbf{b}} = \underset{2\times1}{\mathbf{X'Y}} \tag{6.56}$$

حيث $\mathbf{b}$ متجه معاملات انحدار المربعات الدنيا:

$$\underset{2\times1}{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \tag{6.56a}$$

ولرؤية هذا، تذكّر أننا حصلنا على $\mathbf{X'X}$ في (6.14) و $\mathbf{X'Y}$ في (6.15) وهكذا تعني المعادلة (6.56):

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

أو:

$$\begin{bmatrix} nb_0 + b_1 \sum X_i \\ b_0 \sum X_i + b_1 \sum X_i^2 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

وهذه هي بالضبط المعادلات الناظمية في (6.55)

معاملات الانحدار المقدّرة

لإيجاد معاملات الانحدار المقدّرة من المعادلات الناظمية:

$$\mathbf{X'Xb} = \mathbf{X'Y}$$

بطُرق المصفوفات، نضرب الطرفين عند اليسار بمعكوس $\mathbf{X'X}$ (ونفترضه موجودا):

$$(\mathbf{X'X})^{-1}\mathbf{X'Xb} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

وهكذا نجد، باعتبار أن $(\mathbf{X'X})^{-1}(\mathbf{X'X}) = \mathbf{I}$ و $\mathbf{Ib} = \mathbf{b}$:

$$\underset{2\times1}{\mathbf{b}} = \underset{2\times2}{(\mathbf{X'X})^{-1}}\underset{2\times1}{\mathbf{X'Y}} \tag{6.57}$$

المقدران b_0 و b_1 في **b** هما المقدّران نفساهما المعطيان سابقا في (2.10a) و(2.10b). وسوف نثبت ذلك من خلال مثال.

مثال. لنجد معاملي الانحدار المقدَّرين لمثال حجم الدفعة في شركة وستوود بطرق المصفوفات. فمن عمل سابق لدينا (جدول ٢-٢):

$$n=10 \qquad \sum Y_i = 1{,}100 \qquad \sum X_i = 500 \qquad \sum X_i^2 = 28{,}400$$
$$\sum X_i Y_i = 61{,}800$$

لنستخدم الآن (6.24) لحساب $(\mathbf{X'X})^{-1}$ لدينا:

$$n\sum(X_i-\overline{X})^2 = n\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right] = 10\left[28{,}400 - \frac{(500)^2}{10}\right] = 34{,}000$$

وبالتالي:

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} \dfrac{\sum X_i^2}{n\sum(X_i-\overline{X})^2} & \dfrac{-\sum X_i}{n\sum(X_i-\overline{X})^2} \\ \dfrac{-\sum X_i}{n\sum(X_i-\overline{X})^2} & \dfrac{n}{n\sum(X_i-\overline{X})^2} \end{bmatrix} = \begin{bmatrix} \dfrac{28{,}400}{34{,}000} & \dfrac{-500}{34{,}000} \\ \dfrac{-500}{34{,}000} & \dfrac{10}{34{,}000} \end{bmatrix}$$
$$= \begin{bmatrix} .83529412 & -.01470588 \\ -.01470588 & .00029412 \end{bmatrix}$$

ونرغب أيضا باستخدام (6.15) لحساب $\mathbf{X'Y}$:

$$\mathbf{X'Y} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} 1{,}100 \\ 61{,}800 \end{bmatrix}$$

وبالتالي نجد من (6.57)

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (\mathbf{X'X})^{-1}\mathbf{X'Y} = \begin{bmatrix} .83529412 & -.0147088 \\ -.01470588 & .0029412 \end{bmatrix}\begin{bmatrix} 1{,}100 \\ 61{,}800 \end{bmatrix} = \begin{bmatrix} 10.0 \\ 2.0 \end{bmatrix}$$

أو $b_0 = 10.0$ و $b_1 = 2.0$، ويتفق هذا مع النتائج في الفصل الثاني. وأي اختلاف سيكون بسبب أخطاء التقريب.

ولتقليل تأثير أخطاء التقريب عند إيجاد المتجه **b** بحسابات يدوية، من المستحسن، عادة، إخراج الثابت في مقامات عناصر $(\mathbf{X'X})^{-1}$ خارج المصفوفة، وإجراء القسمة كآخر خطوة. وفي مثالنا سيقود هذا إلى:

$$(\mathbf{X'X})^{-1} = \frac{1}{n\Sigma(X_i-\bar{X})^2}\begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}$$
$$= \frac{1}{34{,}000}\begin{bmatrix} 28{,}400 & -500 \\ -500 & 10 \end{bmatrix}$$
$$\mathbf{b} = \frac{1}{34{,}00}\begin{bmatrix} 28{,}400 & -500 \\ -500 & 10 \end{bmatrix}\begin{bmatrix} 1{,}100 \\ 61{,}800 \end{bmatrix}$$
$$= \frac{1}{34{,}000}\begin{bmatrix} 340{,}000 \\ 68{,}000 \end{bmatrix} = \begin{bmatrix} 10.0 \\ 2.0 \end{bmatrix}$$

وفي حالة كهذه، تقود طريقتا الحساب إلى نتائج متطابقة. على كل حال، فإن تأخير القسمة على $n\Sigma(X_i-\bar{X})^2$ حتى النهاية يعطي في الغالب نتائج أكثر دقة.

تعليقات

١- لاستنباط المعادلتين الناظميتين بطرق المربعات الدنيا، نأخذ القيمة الصغرى للكمية:

$$Q = \Sigma\left[Y_i - (\beta_0 + \beta_1 X_i)\right]^2$$

وبرموز المصفوفة:

$$Q = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

وبفك الأقواس، نجد:

$$Q = \mathbf{Y'Y} - \beta'\mathbf{X'Y} - \mathbf{Y'X}\beta + \beta'\mathbf{X'X}\beta$$

حيث $(\mathbf{X}\beta)' = \beta'\mathbf{X}'$ من (6.33). ونلاحظ الآن أن $\mathbf{Y'X}\beta$ هو 1×1، وبالتالي فهو يساوي منقوله وهو طبقا لـ (6.34) $\beta'\mathbf{X'Y}$ وهكذا نجد:

$$Q = \mathbf{Y'Y} - 2\beta'\mathbf{X'Y} + \beta'\mathbf{X'X}\beta \qquad (6.59)$$

لإيجاد قيمة β التي تجعل Q أصغر مايمكن نشتق بالنسبة لـ β_0 و β_1 ليكن:

$$\frac{\partial}{\partial\beta}(Q) = \begin{bmatrix} \frac{\partial Q}{\partial\beta_0} \\ \frac{\partial Q}{\partial\beta_1} \end{bmatrix} \qquad (6.60)$$

فعندئذ نجد:

$$\frac{\partial}{\partial\beta}(Q) = 2\mathbf{X'Y} + 2\mathbf{X'X}\beta \qquad (6.61)$$

وبمساواتها بالصفر والقسمة على 2 وتعويض **b** عن β نجد الشكل المصفوفي لمعادلات المربعات الدنيا الناظمية:

$$\mathbf{X'Xb = X'Y}$$

٢- تبين مقارنة المعادلات الناظمية مع **X'X** أنه حينما كانت أعمدة **X'X** مرتبطة خطّيا، فستكون المعادلات الناظمية مرتبطة خطّيا أيضا. ولايمكن بالتالي إيجاد حل وحيد لـ b_0 و b_1. ولحسن الحظ، ففي معظم تطبيقات الانحدار تكون أعمدة **X'X** مستقلة خطّيا مما يقود إلى حلين وحيدين لـ b_0 و b_1.

(٦-١١) القيم التوفيقية والرواسب

القيم التوفيقية

ليكن متجه القيم التوفيقية $\hat{\mathbf{Y}}_i$ وسنرمز له بـ $\hat{\mathbf{Y}}$:

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{Y}_n \end{bmatrix} \tag{6.62}$$

وبرموز المصفوفة، لدينا إذن:

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 2}{\mathbf{X}} \ \underset{2\times 1}{\mathbf{b}} \tag{6.63}$$

ذلك لأن:

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ & \cdot \\ & \cdot \\ & \cdot \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \cdot \\ \cdot \\ \cdot \\ b_0 + b_1 X_n \end{bmatrix}$$

مصفوفة قبّعة. يمكننا توضيح نتيجة المصفوفة لـِ $\hat{\mathbf{Y}}$ في (6.63) كما يلي مستخدمين عبارة **b** في (6.57):

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

أو بصورة مكافئة:

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times n}{\mathbf{H}}\ \underset{n\times n}{\mathbf{Y}} \qquad (6.64)$$

حيث:

$$\underset{n\times n}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \qquad (6.64a)$$

وهكذا نرى من (6.64) أنه يمكن التعبير عن القيم التوفيقية $\hat{Y}_i$ كتراكيب خطّية في مشاهدات المتغير التابع Y_i حيث المعاملات هي عناصر المصفوفة **H**. وتنطوي المصفوفة **H** فقط على مشاهدات المتغير المستقل X، كما هو واضح من (6.64a).

تسمى المصفوفة المربعة **H** ذات الأبعاد $n\times n$ ***مصفوفة القبعة***. وتلعب دورا مهما في تحليل الانحدار، كما سترى في الفصل الحادي عشر عندما نناقش ما إذا كانت نتائج الانحدار تتأثر بدون وجه حق بمشاهدة واحدة أو بقليل من المشاهدات. والمصفوفة **H** متناظرة ولها الخاصية (وتدعى تساوي القوى):

$$\mathbf{HH} = \mathbf{H} \qquad (6.65)$$

وبصورة عامة، يقال إن المصفوفة **M** ***متساوية القوة*** إذا كان **MM = M**.

الرواسب

ليكن متجه الرواسب $e_i = Y_i - \hat{Y}_i$ ويرمز له بـ $\boldsymbol{e}$:

$$\underset{n\times 1}{\mathbf{e}} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix} \qquad (6.66)$$

وبرموز المصفوفة، لدينا إذن:

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\mathbf{Xb}} \qquad (6.67)$$

ويمكن التعبير عن الرواسب e_i مثلها مثل القيم التوفيقية $\hat{Y}_i$، كتراكيب خطّيـة في مشاهدات المتغير التابع Y_i وذلك باستخدام النتيجة في (6.64) المتعلقة بـ $\hat{\mathbf{Y}}$:

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

وهكذا نجد النتيجة المهمة:

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times n}{(\mathbf{I} - \mathbf{H})'}\ \underset{n\times 1}{\mathbf{Y}} \qquad (6.68)$$

حيث **H** مصفوفة القبّعة المعرفة في (6.64a).

المصفوفة **I - H** مثلها مثل المصفوفة **H**، متناظرة ومتساوية القوى.

ويمكن تبيان أن مصفوفة تبـاين ـ تغـاير متجه الرواسـب e تنطـوي أيضـا علـى المصفوفة **I - H** :

$$\underset{n\times n}{\sigma^2\{\mathbf{e}\}} = \sigma^2(\mathbf{I} - \mathbf{H}) \qquad (6.69)$$

وتُقدَّر بـ:

$$\underset{n\times n}{s^2\{\mathbf{e}\}} = MSE(\mathbf{I} - \mathbf{H}) \qquad (6.70)$$

ملاحظة

يمكن استنباط مصفوفة تباين ـ تغاير e باستخدام (6.57). فبما أن $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ نجد:

$$\underset{n\times n}{\sigma^2\{\mathbf{e}\}} = (\mathbf{I} - \mathbf{H})\sigma^2\{\mathbf{Y}\}(\mathbf{I} - \mathbf{H})'$$

ولكن $\sigma^2\{\mathbf{Y}\} = \sigma^2\{\varepsilon\} = \sigma^2\mathbf{I}$ في نموذج الخطأ الطبيعي وذلك وفقـا لـ (6.53) ولدينـا أيضا $(\mathbf{I} - \mathbf{H})' = \mathbf{I} - \mathbf{H}$ بسبب تناظر المصفوفة وبالتالي:

$$\begin{aligned}\sigma^2\{\mathbf{e}\} &= \sigma^2(\mathbf{I} - \mathbf{H})\mathbf{I}(\mathbf{I} - \mathbf{H})\\ &= \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\end{aligned}$$

وبالنظر إلى حقيقة كون المصفوفة **I-H** متساوية القوى، نعلم أن:

$$(\mathbf{I}-\mathbf{H})(\mathbf{I}-\mathbf{H}) = (\mathbf{I}-\mathbf{H})$$

ونحصل على العلاقة (6.69):

$$\sigma^2\{\mathbf{e}\} = \sigma^2(\mathbf{I} - \mathbf{H})$$

(٦-١٢) نتائج تحليل التباين

مجموعة المربعات

لرؤية كيفية التعبير عن مجموع المربعات برموز المصفوفة، نبدأ بـ $SSTO$ إذ نعلم من (3.49) أن:

$$SSTO=\sum Y_i^2 - n\bar{Y}^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \qquad (6.71)$$

ونعلم أيضا من (6.13) أن:

$$\mathbf{Y'Y} = \sum Y_i^2$$

يُستخدم الحد المطروح $n\bar{Y}^2 = (\sum Y_i)^2 / n$ معبرا عنه في شكل مصفوفاتي الرمز **J**. وهو مصفوفة المقادير 1 المعرفة في (6.18)، وذلك كما يلي:

$$\frac{(\sum Y_i)^2}{n} = \left(\frac{1}{n}\right)\mathbf{Y'JY} \qquad (6.72)$$

فمثلا، إذا كانت $n = 2$، لدينا:

$$\left(\frac{1}{2}\right)\begin{bmatrix} Y_1 & Y_2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \frac{(Y_1+Y_2)(Y_1+Y_2)}{2}$$

وبالتالي نجد أن:

$$SSTO = \mathbf{Y'Y} - \left(\frac{1}{n}\right)\mathbf{Y'JY} \qquad (6.73a)$$

وتماما كما مثلنا $\sum Y_i^2$ بـ $\mathbf{Y'Y}$ بدلالة المصفوفات، كذلك يمكن تمثيل $SSE = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$ كالتالي:

$$SSE = \mathbf{e'e} = (\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb}) \qquad (6.73b)$$

ويمكن تبيان أنه يساوي:

$$SSE = \mathbf{Y'Y} - \mathbf{b'X'Y} \qquad (6.73c)$$

وأخيرا يمكن تبيان أن:

$$SSR = \mathbf{b'X'Y} - \left(\frac{1}{n}\right)\mathbf{Y'JY} \qquad (6.73d)$$

مثال. دعنا نجد SSE لمثال حجم الدفعة في شركة وستوود بطرق المصفوفة، مستخدمين (6.73c). فنعلم من نتائج سابقة أن:

$$\mathbf{Y'Y} = \sum Y_i^2 = 134{,}660$$

ونعلم أيضا من نتائج سابقة أن:

$$\mathbf{b} = \begin{bmatrix} 10.0 \\ 2.0 \end{bmatrix} \qquad \mathbf{X'Y} = \begin{bmatrix} 1{,}100 \\ 61{,}800 \end{bmatrix}$$

وبالتالي:

$$\mathbf{b'X'Y} = \begin{bmatrix} 10.0 & 2.00 \end{bmatrix} \begin{bmatrix} 1{,}100 \\ 61{,}800 \end{bmatrix} = 134{,}600$$

و:

$$SSE = \mathbf{Y'Y} - \mathbf{b'X'Y} == 134{,}660 - 134{,}600 = 60$$

وهي النتيجة نفسها التي حصلنا عليها في الفصل الثاني وأية فروق ستكون بسبب أخطاء التقريب.

وبصورة مماثلة يمكننا إيجاد SSE باستخدام (6.73d):

$$SSR = \mathbf{b'X'Y} - \left(\frac{1}{n}\right) \mathbf{Y'JY}$$

$$= 134{,}600 - \frac{1}{10}(1{,}100)^2 = 13{,}600$$

ذلك لأن $\sum Y_i = 1{,}100$ لمثال شركة وستوود.

ملاحظة

لإيضاح استنباط تعابير مجاميع المربعات برموز المصفوفة، خذ SSE:

$$SSE = \mathbf{e'e} = (\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb}) = \mathbf{Y'Y} - 2\mathbf{b'X'Y} + \mathbf{b'X'Xb}$$

وبالتعويض عن b في أقصى اليمين نجد من (6.57):

$$SSE = \mathbf{Y'Y} - 2\mathbf{b'X'Y} + \mathbf{b'X'X(X'X)^{-1}X'Y}$$
$$= \mathbf{Y'Y} - 2\mathbf{b'X'Y} + \mathbf{b'IX'Y}$$

وبحذف $\mathbf{I}$ والطرح، نجد النتيجتين في (6.73c):

$$SSE = \mathbf{Y'Y} - \mathbf{b'X'Y}$$

مجاميع المربعات كصيغ تربيعية

يمكن تبيان أن مجاميع مربعات هي صيغ تربيعية، كمثال لصيغة تربيعية في المشاهدات Y_i عندما $n = 2$ نأخذ:

$$5Y_1^2 + 6Y_1Y_2 + 4Y_2^2 \qquad (6.74)$$

لاحظ أن هذه العبارة هي كثيرة حدود من الدرجة الثانية تحتوي على حدود تتضمن مربعات المشاهدات وجداءاتها ونعبر عن (6.74) بدلالة المصفوفات كما يلي:

$$\begin{bmatrix} Y_1 & Y_2 \end{bmatrix} \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{Y'AY} \qquad (6.74a)$$

حيث **A** مصفوفة متناظرة.

وبصورة عامة، نعرف صيغة تربيعية كما يلي:

$$\underset{1\times1}{\mathbf{Y'AY}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}Y_iY_j \qquad \text{حيث : } a_{ij} = a_{ji} \qquad (6.75)$$

A مصفوفة $n\times n$ متناظرة وتسمى *مصفوفة الصيغة التربيعية*.

ومجاميع مربعات التحاين *SSTO*، *SSE* و*SSR* جميعها صيغ تربيعية. ولرؤية ذلك، نحتاج إلى التعبير بصورة أكثر تراصا عن الأشكال المصفوفية لمجاميع المربعات هذه المذكورة في (6.73) ونقوم بذلك بإعادة التعبير عن $\mathbf{b'X'}$. فمن (6.33) و(6.63)، نعلم أن:

$$\mathbf{b'X'} = (\mathbf{Xb})' = \hat{\mathbf{Y}}'$$

الآن نستخدم النتيجة في (6.64) للحصول على:

$$\mathbf{b'X'} = (\mathbf{HY})'$$

وبما أن **H** مصفوفة متناظرة فلدينا $\mathbf{H'} = \mathbf{H}$، ونجد أخيرا، باستخدام (6.33):

$$\mathbf{b'X'} = \mathbf{Y'H} \qquad (6.76)$$

وتسمح هذه النتيجة بالتعبير عن مجاميع المربعات في (6.73) كما يلي:

$$SSTO = \mathbf{Y'}\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \qquad (6.77a)$$

$$SSE = \mathbf{Y'}(\mathbf{I} - \mathbf{H})\mathbf{Y} \qquad (6.77b)$$

$$SSR = \mathbf{Y'}\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \qquad (6.77c)$$

يمكن الآن رؤية أن كلا من مجاميع المربعات هذه هو من الشكل **Y'AY** ويمكن تبيان أن المصفوفات **A** الثلاث:

$$\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J} \tag{6.78a}$$

$$\mathbf{I}-\mathbf{H} \tag{6.78b}$$

$$\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J} \tag{6.78c}$$

هي مصفوفات متناظرة. وبالتالي *SSTO*، *SSE* و*SSR* هي صيغ تربيعية، بمصفوفات صيغ تربيعية معطاة في (6.78). وتلعب الصيغ التربيعية دورا مهما في الإحصاء لأن كل مجاميع المربعات في تحليل التباين للنماذج الإحصائية الخطّية يمكن التعبير عنها كصيغ تربيعية.

(٦-١٣) استقراءات في تحليل الانحدار

كما رأينا في فصول سابقة، جميع التقديرات بفترة هي من الشكل التالي: مقدِّر نقطي زائد أو ناقص عددا معينا من الانحرافات المعيارية المقدَّرة للمقدِّر النقطي. وبصورة مماثلة تتطلب جميع الاختبارات مقدَّرا نقطيا والانحراف المعياري المقدَّر للمقدِّر النقطي أو في حالة اختبارات تحليل تباين، مجاميع المربعات المختلفة. ولجبر المصفوفات فائدة رئيسة في الاستقراء وذلك عند الحصول على تقديرات الانحرافات المعيارية وعلى مجاميع المربعات. ولقد اعطينا آنفا المكافئات المصفوفية لمجاميع المربعات لتحليل التباين. وبالتالي، نركز هنا بصورة رئيسة على التعابير المصفوفية لتقديرات تباينات المقدِّرات النقطية موضع الاهتمام.

معاملات انحدار

مصفوفة تباين - تغاير **b** التالية:

$$\underset{2\times 2}{\sigma^2\{\mathbf{b}\}} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0,b_1\} \\ \sigma\{b_1,b_0\} & \sigma^2\{b_1\} \end{bmatrix} \tag{6.79}$$

هي:

$$\underset{2\times 2}{\sigma^2\{\mathbf{b}\}} = \sigma^2\{\mathbf{X'X}\}^{-1} \tag{6.80}$$

أو باستخدام (6.25)

$$\underset{2\times2}{\sigma^2\{\mathbf{b}\}}=\begin{bmatrix}\dfrac{\sigma^2\sum X_i^2}{n\sum(X_i-\overline{X})^2} & \dfrac{-\overline{X}\sigma^2}{\sum(X_i-\overline{X})^2}\\ \dfrac{-\overline{X}\sigma^2}{\sum(X_i-\overline{X})^2} & \dfrac{\sigma^2}{\sum(X_i-\overline{X})^2}\end{bmatrix} \tag{6.80a}$$

عند التعويض بـ MSE عن σ^2 في (6.80)، نجد:

$$\underset{2\times2}{\mathbf{s}^2\{\mathbf{b}\}}=MSE(\mathbf{X'X})^{-1}=\begin{bmatrix}\dfrac{MSE\sum X_i^2}{n\sum(X_i-\overline{X})^2} & \dfrac{-\overline{X}MSE}{\sum(X_i-\overline{X})^2}\\ \dfrac{-\overline{X}MSE}{\sum(X_i-\overline{X})^2} & \dfrac{MSE}{\sum(X_i-\overline{X})^2}\end{bmatrix} \tag{6.81}$$

حيث $\mathbf{s}^2\{\mathbf{b}\}$ هي مصفوفة تباين ـ تغاير $\mathbf{b}$ المقدّرة. وفي $(6.80a)$ سنتعرف على تباين b_0 في $(3.20b)$ و تباين b_1 في (3.3b) وتغاير b_0 وb_1 في (5.6). وبصورة مشابهة، فإن التباينات المقدرة في (6.81)، هي تباينات مألوفة من الفصول السابقة.

متوسط استجابة

لتقدير متوسط الاستجابة عند $\mathbf{X}_h$، دعنا نعرّف المتجه:

$$\underset{2\times1}{\mathbf{X}_h}=\begin{bmatrix}1\\X_h\end{bmatrix} \quad \text{او} \quad \underset{1\times2}{\mathbf{X}'_h}=[1\ \ X_h] \tag{6.82}$$

فعندئذ ستكون القيمة التوفيقية برموز المصفوفة كما يلي:

$$\hat{Y}_h=\mathbf{X}'_h\,\mathbf{b} \tag{6.83}$$

ذلك لأن:

$$\mathbf{X}'_h\,\mathbf{b}=\begin{bmatrix}1 & X_h\end{bmatrix}\begin{bmatrix}b_0\\b_1\end{bmatrix}=\begin{bmatrix}b_0+b_1X_h\end{bmatrix}=\begin{bmatrix}\hat{Y}_h\end{bmatrix}=\hat{Y}_h$$

ونلاحظ أن $\mathbf{X}'_h\,\mathbf{b}$ مصفوفة 1×1؛ وبالتالي، يمكن كتابة النتيجة النهائية كعدد سلّمي.

وتباين $\hat{Y}_h$ المعطى سابقا في $(3.28b)$ يصبح برموز المصفوفة:

$$\sigma^2\{\hat{\mathbf{Y}}_h\}=\sigma^2\left(\mathbf{X}'_h(\mathbf{X'X})^{-1}\mathbf{X}_h\right)=\mathbf{X}'_h\sigma^2\{\mathbf{b}\}\mathbf{X}_h \tag{6.84}$$

حيث $\sigma^2\{\mathbf{b}\}$ مصفوفة تباين ـ تغاير معاملات الانحدار في (6.80) ونلاحظ بالتالي، أن

$\sigma^2\{\hat{Y}_h\}$ دالة في التباينين $\sigma^2\{b_0\}$ و$\sigma^2\{b_1\}$ والتغاير $\sigma\{b_0, b_1\}$.

والتباين المقدر لـ $\hat{Y}_h$. المعطى سابقا في (3.29) يصبح برموز المصفوفة:

$$s^2\{\hat{Y}_h\} = MSE\left(\mathbf{X}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right) = \mathbf{X}'_h s^2\{\mathbf{b}\}\mathbf{X}_h \qquad (6.85)$$

حيث $s^2\{b\}$ مصفوفة التباين ـ التغاير المقدَّرة لمعاملات الانحدار في (6.81)

التنبؤ بمشاهدة جديدة

التباين المقدَّر $s^2\{Y_{h(\text{new})}\}$ ، المعطى سابقا في (3.37) يصبح برموز المصفوفة:

$$s^2\{Y_{h(\text{new})}\} = MSE + s^2\{\hat{Y}_h\} = MSE + \mathbf{X}'_h s^2\{\mathbf{b}\}\mathbf{X}_h \qquad (6.86)$$

$$= MSE\left(1 + X'_h(X'X)^{-1}X_h\right)$$

أمثلة

١ـ نرغب ايجاد $s^2\{b_0\}$ و$s^2\{b_1\}$ لمثال حجم الدفعة في شركة وستوود بطرق المصفوفة. وجدنا سابقا أن $MSE = 7.5$ و:

$$(X'X)^{-1} = \begin{bmatrix} .83529412 & -.01470588 \\ -.01470588 & .00029412 \end{bmatrix}$$

ومن (6.81) نجد التالي:

$$s^2\{b\} = MSE(X'X)^{-1} = 7.5\begin{bmatrix} .83529412 & -.01470588 \\ -.01470588 & .00029412 \end{bmatrix}$$

$$= \begin{bmatrix} 6.264706 & -.1102941 \\ -.1102941 & .0022059 \end{bmatrix}$$

وهكذا $s^2\{b_0\} = 6.26471$ و$s^2\{b_1\}=0.002206$ وهي النتائج نفسها التي حصلنا عليها في الفصل الثالث.

لاحظ مدى بساطة ايجاد تبايني معاملي الانحدار حالما نجد $(\mathbf{X}'\mathbf{X})^{-1}$. ونحتاج هذا المعكوس في المقام الأول لإيجاد معاملي الانحدار، وهكذا فلا يتطلب ايجاد تباينيهما المقدرين، عمليا، أية جهود إضافية.

٢ـ نرغب إيجاد $s^2\{\hat{Y}_h\}$ لمثال شركة وستوود عندما $\mathbf{X}_h = 55$ نعرف:

$$X'_h = [1 \quad 55]$$

$$X'_h = [1 \quad 55]$$

ونجد من (6.85) أن:

$$s^2\{\hat{Y}_n\} = \mathbf{X}'_h s^2 \{\mathbf{b}\} \mathbf{X}_h$$
$$= [1 \quad 55]\begin{bmatrix} 6.264706 & -.1102941 \\ -.1102941 & .0022059 \end{bmatrix}\begin{bmatrix} 1 \\ 55 \end{bmatrix} = 0.80520$$

وهذه هي النتيجة نفسها التي حصلنا عليها في الفصل الثالث، باستثناء فرق بسيط يعود إلى تدوير الأرقام العشرية.

تعليقات

١- لتوضيح عملية استنباط بدلالة المصفوفات، دعنا نجد مصفوفة تبـاين ــ تغـاير **b**. تذكّر أن:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{AY}$$

حيث **A** مصفوفة ثابتة:

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

وبالتالي لدينا من (6.47):

$$\sigma^2\{\mathbf{b}\} = \mathbf{A}\sigma^2\{\mathbf{Y}\}\mathbf{A}'$$

والآن $\sigma^2\{\mathbf{Y}\} = \sigma^2\mathbf{I}$ وبالإضافة إلى ذلك نجد من (6.33)، ومـن حقيقـة أن $(\mathbf{X}'\mathbf{X})^{-1}$ متناظرة، أن :

$$\mathbf{A}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

ولذلك نجد:

$$\sigma^2\{\mathbf{b}\} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$
$$= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$
$$= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{I}$$
$$= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

٢- وبما أن $\hat{Y}_n = \mathbf{X}'_h\mathbf{b}$ ، فنستنتج من (6.47) على الفور، أن:

$$\sigma^2\{\hat{Y}_h\} = \mathbf{X}'_h\,\sigma^2\{\mathbf{b}\}\mathbf{X}_h$$

وبالتالي:

$$\sigma^2\{\hat{Y}_h\}=\begin{bmatrix}1 & X_h\end{bmatrix}\begin{bmatrix}\sigma^2\{b_0\} & \sigma\{b_0,b_1\}\\ \sigma\{b_1,b_0\} & \sigma^2\{b_0\}\end{bmatrix}\begin{bmatrix}1\\ X_h\end{bmatrix}$$

أو:

$$\sigma^2\{\hat{Y}_h\}=\sigma^2\{b_0\}+2X_h\sigma\{b_0,b_1\}+X_h^2\sigma^2\{b_1\} \qquad (6.87)$$

وباستخدام النتائج من $(6.80a)$ نجد:

$$\sigma^2\{\hat{Y}_h\}=\frac{\sigma^2\sum X_i^2}{n\sum(X_i-\overline{X})^2}+\frac{2X_h(-\overline{X})\sigma^2}{\sum(X_i-\overline{X})^2}+\frac{X_h^2\sigma^2}{\sum(X_i-\overline{X})^2}$$

والتي تختزل إلى العبارة المألوفة:

$$\sigma^2\{\hat{Y}_h\}=\sigma^2\left[\frac{1}{n}+\frac{(X_h-\overline{X})^2}{\sum(X_i-\overline{X})^2}\right] \qquad (6.88)$$

وهكذا نرى صراحة أن تعبير التباين في (6.88) يحوي مساهمات من $\sigma^2\{b_0\}$، $\sigma^2\{b_1\}$ و $\sigma\{b_0,b_1\}$ ويجب أن يكون طبقا للنظرية $(1.27b)$:

$$\hat{Y}_h=b_0+b_1X_h$$

باعتبار أن $\hat{Y}_h$ تركيب خطّي في b_0 و b_1.

٣- لا نعرض النتائج بدلالة المصفوفة لبقية أنواع الاستقراءات، مثل التنبؤ المتزامن لعدد من المشاهدات Y الجديدة، عند مستويات مختلفة لـ X_h، ذلك لأنها تستند إلى نتائج قمنا بتطويرها.

مراجع ورد ذكرها

[6.1] Graybill, F. A. *Matrices with Applications in Statistics. 2nd ed.* Belmont, Calif. : Wadsworth, 1983.

مسائل

(٦-١) أوجد للمصفوفات المذكورة أدناه: (١) **A+B**، (٢) **A-B**، (٣) **AC**، (٤) **AB′** و(٥) **B′A**

$$\mathbf{A}=\begin{bmatrix}1 & 4\\ 2 & 6\\ 3 & 8\end{bmatrix}\quad \mathbf{B}=\begin{bmatrix}1 & 3\\ 1 & 4\\ 2 & 5\end{bmatrix}\quad \mathbf{C}=\begin{bmatrix}3 & 8 & 1\\ 5 & 4 & 0\end{bmatrix}$$

اذكر البعد لكل مصفوفة ناتجة.

(٦-٢) أوجد للمصفوفات أدناه: (١) A+C، (٢) A-C، (٣) B'A، (٤) AC' و(٥) C' A.

$$\mathbf{A}=\begin{bmatrix}2&1\\3&5\\5&7\\4&8\end{bmatrix}\quad \mathbf{B}=\begin{bmatrix}6\\9\\3\\1\end{bmatrix}\quad .\ \mathbf{C}=\begin{bmatrix}3&8\\8&6\\5&1\\2&4\end{bmatrix}$$

اذكر البعد لكل مصفوفة ناتجة.

(٦-٣) بيّن كيفية كتابة العبارات التالية بدلالة المصفوفات: (١) $Y_i-\hat{Y}_i=e_i$ و(٢) $\sum X_i\, e_i=0$ افترض $i=1,...,4$.

(٦-٤) **فساد النكهة.** تم الحصول على النتائج المبينة أدناه من تجربة على نطاق ضيق لدراسة العلاقة بين درجة التخزين X بالفهرنهايت وعدد الأسابيع قبل بدء فساد النكهة Y.

i	1	2	3	4	5
X_i	8	4	0	-4	-8
Y_i	7.8	9.0	10.2	11.0	11.7

افترض أن نموذج الانحدار من المرتبة – الأولى (3.1) قابل للتطبيق. أوجد باستخدام طرق المصفوفة.

(١) $\mathbf{Y'Y}$ (٢) $\mathbf{X'X}$ (٣) $\mathbf{X'Y}$

(٦-٥) **تمويل مستهلك.** تعرض البيانات أدناه لشركة تمويل المستهلك تعمل في ست مدن، عدد شركات الإقراض المنافسة العاملة في المدينة (X) وعدد القروض المستحقة التي لم تُسدد للشركة في تلك المدينة بالآلاف (Y).

i	1	2	3	4	5	6
X_i	4	1	2	3	3	4
Y_i	16	5	10	15	13	22

افترض أن نموذج الانحدار من المرتبة الأولى (3.1) قابل للتطبيق. مستخدما طرق المصفوفة، أوجد : (١) $\mathbf{Y'Y}$ (٢) $\mathbf{X'X}$ (٣) $\mathbf{X'Y}$

(٦-٦) عد إلى مسألة **تكسر الشحنات** (٢-١٩) مستخدما طرق المصفوفة، أوجد:

$X'Y$ (٣)　　$X'X$ (٢)　　$Y'Y$ (١)

(٦-٧) عُد إلى مسألة **صلابة البلاستيك** (٢-٢٠) مستخدما طرق المصفوفة، أوجد:

$X'Y$ (٣)　　$X'X$ (٢)　　$Y'Y$ (١)

(٦-٨) لتكن **B** معرّفة كالتالي:

$$\mathbf{B}=\begin{bmatrix}1 & 5 & 0\\ 1 & 0 & 5\\ 1 & 0 & 5\end{bmatrix}$$

ا - هل متجهات العمود في **B** مرتبطة خطّيا؟

ب - ما هي رتبة **B**؟

جـ - ماذا يجب أن تكون محدّدة **B**؟

(٦-٩) لتكن **A** معرّفة كالتالي:

$$\mathbf{A}=\begin{bmatrix}0 & 1 & 8\\ 0 & 3 & 1\\ 0 & 5 & 5\end{bmatrix}$$

ا - هل متجهات العمود في **A** مرتبطة خطّيا؟

ب - أعد صياغة تعريف (6.20) بدلالة متجهات الصف. هل متجهات الصف في **A** مرتبطة خطّيا ؟

جـ - ما هي رتبة **A**؟

د - احسب محدّدة **A**.

(٦-١٠) أوجد المعكوس لكل من المصفوفتين التاليتين:

$$\mathbf{A}=\begin{bmatrix}2 & 4\\ 3 & 1\end{bmatrix} \quad \mathbf{B}=\begin{bmatrix}4 & 3 & 2\\ 6 & 5 & 10\\ 10 & 1 & 6\end{bmatrix}$$

تأكد في كل حالة أن المصفوفة الناتجة هي بالفعل المعكوس.

(٦-١١) أوجد المعكوس للمصفوفة التالية:

$$\mathbf{A} = \begin{bmatrix} 5 & 1 & 3 \\ 4 & 0 & 5 \\ 1 & 9 & 6 \end{bmatrix}$$

تأكد أن المصفوفة الناتجة هي بالفعل المعكوس.

(٦-١٢) عُد إلى مسألة **فساد النكهة** (٦-٤). أوجد $(\mathbf{X'X})^{-1}$

(٦-١٣) عُد إلى مسألة **تمويل المستهلك** (٦-٥). أوجد $(\mathbf{X'X})^{-1}$

(٦-١٤) اعتبر المعادلتين الآتيتين:

$$4y_1 + 7y_2 = 25$$
$$2y_1 + 3y_2 = 12$$

ا ـ اكتب هاتين المعادلتين برموز المصفوفة.

ب ـ مستخدما طرق المصفوفة، أوجد الحلول لـy_1 وy_2.

(٦-١٥) اعتبر المعادلتي المتزامنتين:

$$5y_1 + 2y_2 = 8$$
$$23y_1 + 7y_2 = 28$$

ا ـ اكتب هاتين المعادلتين برموز المصفوفة.

ب ـ مستخدما طرق المصفوفة، أوجد الحلول لـ y_2 وy_1.

(٦-١٦) اعتبر دالة الانحدار الخطية المقدَّرة في شكلها المعطى في (2.15). اكتب تعابير القيم التوفيقية $\hat{Y}_i$ كما تعطيها (2.15) من أجل $i = 1,...,5$ وذلك بدلالة المصفوفات.

(٦-١٧) اعتبر الدوال التالية في المتغيرات العشوائية Y_1، Y_2 و Y_3 :

$$W_1 = Y_1 + Y_2 + Y_3$$
$$W_2 = Y_1 - Y_2$$
$$W_3 = Y_1 - Y_2 - Y_3$$

ا ـ اعرض المعادلات أعلاه برموز المصفوفات.

ب ـ أوجد توقع المتجه العشوائي $\mathbf{W}$

جـ ـ أوجد مصفوفة تباين ـ تغاير $\mathbf{W}$

(٦-١٨) اعتبر الدالتين التاليتين في المتغيرات العشوائية Y_1، Y_2، Y_3 و Y_4

$$W_1 = \frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)$$

$$W_2 = \frac{1}{2}(Y_1 + Y_2) - \frac{1}{2}(Y_3 + Y_4)$$

ا - اعرض المعادلتين أعلاه برموز المصفوفة.

ب - أوجد توقع المتجه العشوائي **W**.

جـ - أوجد مصفوفة تباين - تغاير لـ**W**.

(٦-١٩) أوجد المصفوفة **A** للصيغة التربيعية:

$$3Y_1^2 + 10Y_1Y_2 + 17Y_2^2$$

(٦-٢٠) أوجد المصفوفة **A** للصيغة التربيعية:

$$7Y_1^2 - 8Y_1Y_2 + 8Y_2^2$$

(٦-٢١) من أجل المصفوفة:

$$\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$$

أوجد الصيغة التربيعية للمشاهدتين Y_1 و Y_2.

(٦-٢٢) من أجل المصفوفة:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 3 & 0 \\ 4 & 0 & 9 \end{bmatrix}$$

أوجد الصيغة التربيعية للمشاهدات Y_1، Y_2 و Y_3.

(٦-٢٣) بالعودة إلى مسألتي **فساد النكهة** (٦-٤) و(٦-١٢).

ا - مستخدما طرق المصفوفة، أوجد التالي: (١) متجه معاملات الانحدار المقدرة، (٢) متجه الرواسب، (٣) SSR، (٤) SSE، (٥) مصفوفة تباين - تغاير **b** المقدرة، (٦) تقدير نقطي لِـ $E\{Y_h\}$ عندما يكون $\mathbf{X}_h = -6$ و(٧) تباين $\hat{Y}_h$ المقدّر عند $\mathbf{X}_h = -6$.

ب - ما هي التبسيطات التي تنشأ عن تحديد المسافات الفاصلة بين مستويات X في التجربة؟

جـ - أوجد مصفوفة القبّعة **H**.

د - أوجد $s^2\{e\}$.

(٦-٢٤) بالإشارة إلى مسألتي **تمويل مستهلك** (٦-٥) و(٦-١٣).

ا - أوجد ما يلي مستخدما طرق المصفوفات: (١) متجه معاملات الانحدار المقدَّرة، (٢) متجه الرواسب، (٣) SSR، (٤) SSE، (٥) مصفوفة التغاير المقدَّرة للمتجه $\mathbf{b}$، (٦) تقديرا نقطيا لِ $E\{Y_h\}$ عندما يكون $X_h = 4$ و (٧) التباين المقدَّر لِ $Y_{h(new)}$ عندما يكون $X_h = 4$.

ب - من مصفوفة التباين - التغاير المقدرة في السؤال (أ) (٥)، أوجد التالي:

(١) $s\{b_0, b_1\}$، (٢) $s^2\{b_0\}$ و (٣) $s\{b_1\}$

جـ - أوجد مصفوفة القبعة $\mathbf{H}$.

د - أوجد $s^2\{e\}$.

(٦-٢٥) عُد إلى مسألتي **تكسر الشحنات** (٢-١٩) و(٦-٦).

ا - مستخدما طرق المصفوفة، أوجد التالي:

(١) $(\mathbf{X'X})^{-1}$، (٢) $\mathbf{b}$، (٣) $\mathbf{e}$، (٤) $\mathbf{H}$، (٥) SSE، (٦) $s^2\{\mathbf{b}\}$، (٧) $\hat{Y}_h$ عندما $X_h = 2$ و(٨) $s^2\{\hat{Y}_h\}$ عندما $X_h = 2$.

ب - من السؤال (أ) (٦) أوجد التالي:

(١) $s^2\{b_1\}$ (٢) $s\{b_0, b_1\}$ و(٣) $s\{b_0\}$

جـ - أوجد مصفوفة الصيغة التربيعية لِ SSR

(٦-٢٦) عد إلى مسألتي **صلابة البلاستيك** (٢-٢٠) و(٦-٧).

ا - مستخدما طرق المصفوفة، أوجد التالي:

(١) $(\mathbf{X'X})^{-1}$، (٢) $\mathbf{b}$، (٣) $\hat{\mathbf{Y}}$، (٤) $\mathbf{H}$، (٥) SSE، (٦) $s^2\{\mathbf{b}\}$ و (٧) $s^2\{Y_{h(new)}\}$ عندما $X_h = 30$

ب - من السؤال (أ-٦)، أوجد التالي:

(١) $s^2\{b_0\}$، (٢) $s\{b_0, b_1\}$ (٣) $s\{b_1\}$

جـ - أوجد مصفوفة الصيغة التربيعية لـ SSE

تمارين

(٦-٢٧) بالعودة إلى نموذج الانحدار عبر نقطة الأصل (5.11) اكتب متجه التوقع لـ ε. افترض أن $i = 1,..., 4$.

(٦-٢٨) اعتبر نموذج (5.11) للانحدار عبر الأصل والمقدّر b_1 المعطى في (5.15) أوجد (5.15) مستفيدا من (6.57) ومعرفا **X** بصورة مناسبة.

(٦-٢٩) اعتبر مقدّر المربعات الدنيا **b** المعطى في (6.57) مستخدما طرق المصفوفة، بيّن أن **b** مقدّر غير منحاز.

(٦-٣٠) بيّن أنه يمكن التعبير عن $\hat{Y}_h$ في (6.83) بدلالة المصفوفات على الشكل $\mathbf{b'X_h}$.

(٦-٣١) أوجد عبارة مصفوفة تباين - تغاير القيم التوفيقية $\hat{Y}_i$، $i = 1,...,n$ بدلالة مصفوفة القبعة.

الباب الثاني

الانحدار الخطي العام

- الانحدار المتعدد – I
- الانحدار المتعدد – II
- انحدار كثيرات الحدود
- المتغيرات المستقلة النوعية
- تشخيصات وتدابير علاجية – II
- بناء نموذج الانحدار
- الارتباط الذاتي في بيانات السلاسل الزمنية



الانحدار المتعدد ـ I

تحليل الانحدار المتعدد هو، من بين كافة الأدوات الإحصائية، الأداة المستخدمة على أوسع نطاق. وفي هذا الفصل، سنناقش أولا تشكيلة من نماذج الانحدار المتعدد. ثم نقدم النتائج الإحصائية الأساسية للانحدار المتعدد في صيغة مصفوفية. وبما أن التعابير المصفوفية للانحدار المتعدد هي نفسها كما في الانحدار الخطي البسيط فسنعرض النتائج دون كثير من النقاش. ثم نعطي مثالا يوضح تشكيلة من الاستدلالات (الاستقراءات) وتحليلات الرواسب في تحليل الانحدار المتعدد.

(٧-١) نماذج الانحدار المتعدد

الحاجة لعدة متغيرات مستقلة

عندما قدمنا تحليل الانحدار، للمرة الأولى، في الفصل الثاني، تحدثنا عن نماذج انحدار تتضمن عددا من المتغيرات المستقلة. وقد ذكرنا نموذج انحدار كان المتغير التابع فيه هو كلفة التشغيل المباشر لمكتب فرعي في سلسلة من مكاتب تمويل العملاء، وأخذنا في الاعتبار أربعة متغيرات مستقلة تضمنت متوسط عدد القروض غير المدفوعة في الفرع، والعدد الكلي لطلبات القروض الجديدة التي يقوم الفرع بمعالجتها. وذكرنا أيضا دراسة تتعلق بشراء محراث آلي كان متغير الاستجابة فيها هو حجم مشتروات المحاريث الآلية في منطقة بيع، وتضمنت المتغيرات المستقلة التسعة عدد المزارع في المنطقة، ومقدار إنتاج المحاصيل في المنطقة. وبالاضافة إلى ذلك ذكرنا دراسة عن الأطفال القصار حيث كان متغير الاستجابة فيها هو أعلى مستوى لهرمون نمو البلازما، وتضمنت المتغيرات المستقلة الأربعة عشر الجنس والعمر وقياسات مختلفة في الجسم. وفي جميع هذه الأمثلة، سوف لا يزوّدنا متغير مستقل واحد بوصف ملائم طالما أن عددا من المتغيرات المستقلة الرئيسة تؤثر بطرق مهمة ومتميزة في متغير الاستجابة. وفضلا عن ذلك، فكثيرا ماسيجد المرء في حالات من هذا النوع، أن التنبؤات بقيم متغير الاستجابة المستندة إلى نموذج يتضمن

متغيرا مستقلا واحدا فقط، هي من عدم الدقة بحيث تصبح عديمة الفائدة. والنموذج الأكثر تعقيدا المتضمن لمتغيرات مستقلة إضافية، هو عادة أكثر عونا في تقديم تنبؤات دقيقة بكفاية لمتغير الاستجابة.

وفي كل من الأمثلة السابقة استند التحليل على بيانات مشاهدة لأن بعضا من المتغيرات المستقلة أو جميعها لم يكن من المتغيرات التي يمكن التحكم فيها مباشرة. وعندما يكون المجرّب قادرا على التحكّم بالمتغيرات المستقلة يكون تحليل الانحدار المتعدد جَمَّ الفائدة أيضا. وسيرغب المجرّب، تقليديا ، بتقصّي عدد من المتغيرات المستقلة في آن واحد، إذ دائما ما يؤثر أكثر من متغير رئيس مستقل واحد في الاستجابة. فعلى سبيل المثال، في دراسة لإنتاجية طاقم عمل، قد يرغب المجرب في التحكم في كل من حجم الطاقم ومستوى العلاوات المدفوعة. وبصورة مماثلة، في دراسة الاستجابة لدواء، قد يرغب المجرب في التحكم في كل من مقدار الجرعة وطريقة تناولها.

ويمكن الاستفادة من نماذج الانحدار المتعدد التي سنصفها الآن في كل من بيانات المشاهدة، والبيانات التجريبية الناتجة عن تصميم تام العشوائية.

نموذج من المرتبة الأولى مع متغيرين مستقلين

عندما يوجد متغيران مستقلان X_1 و X_2 يُدعى نموذج الانحدار:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad (7.1)$$

نموذجا من المرتبة الأولى مع متغيرين مستقلين. وكما نذكر من الفصل الثاني فإن نموذجا من المرتبة الأولى هو نموذج خطّي في المعالم وخطّي في المتغيرات المستقلة. ويرمز Y_i كالمعتاد للاستجابة في التكرار i ، X_{i1} و X_{i2} هما قيمتا المتغيرين المستقلين في التكرار i. ومعالم النموذج هي β_0 ، β_1 و β_2، و ε_i هو حدّ الخطأ.

وبافتراض $E\{\varepsilon_i\} = 0$ تكون دالة انحدار النموذج (7.1) هي:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (7.2)$$

وبصورة مشابهة للانحدار الخطّي البسيط، حيث تكون دالة الانحدار $E\{Y\} = \beta_0 + \beta_1 X$ خطا مستقيما، فإن دالة الانحدار (7.2) هي مستوي. ويتضمن الشكل (٧-١) تمثيلا لجزء من مستوي الاستجابة:

$$E\{Y\} = 20.0 + 0.95X_1 - 0.50X_2 \qquad (7.3)$$

نلاحظ أن نقطة على مستوي الاستجابة (7.3) تقابل متوسط الاستجابة $E\{Y\}$ عند

مركّب معطى لمستويي X_1 و X_2.

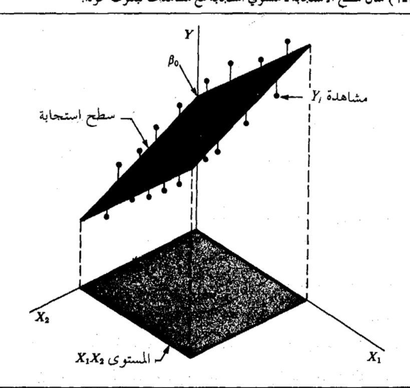
ويبين الشكل (٧–١) أيضا سلسلة من المشاهدات Y_iعلى مستوي الاستجابة، الموافقة لمستويين محددين (X_{i1}, X_{i2}) للمتغيرين المستقلين. ونلاحظ أن الخطوط الرأسية في الشكل (٧–١) تمثل الفرق بين Y_i والمتوسط $E\{Y_i\}$ الواقع على مستوي الاستجابة للتوزيع الاحتمالي لِ (X_{i1}, X_{i2}) وبالتالي تمثل المسافة الرأسية بين Y_i ومستوي الاستجابة حد الخطأ $\varepsilon_i = Y_i - E\{Y_i\}$.

وكثيرا ما تُدعى دالة الانحدار في الانحدار المتعدد، سطح الانحدار أو سطح الاستجابة. وفي الشكل (٧–١)، نجد أن سطح الاستجابة هو مجرد مستوي، إلا أن سطح الاستجابة في حالات أخرى يمكن أن يكون معقدا في طبيعته.

معنى معاملات الانحدار. لنعتبر الآن معنى معاملات الانحدار في دالة الانحدار المتعدد (7.3). فالمعلمة $\beta_0 = 20.0$ هي تقاطع Y مع مستوي الانحدار. وإذا امتد مجال النموذج ليشمل $X_1 = 0$ و $X_2 = 0$ فإن $\beta_0 = 20.0$ تعطي متوسط الاستجابة عند $X_1 = 0$ و $X_2 = 0$.

وفيما عدا ذلك لايكون لِ β_0 كحد منفصل في نموذج الانحدار، أي معنى محدّد.

شكل (٧-١) مثال سطح الاستجابة ـ مستوي استجابة مع مشاهدات تبعثرت حوله.



وتشير المعلمة β_1 إلى التغيّر في متوسط الاستجابة لكل زيادة بمقدار الواحد في X_1 وذلك عندما يبقى X_2 ثابتا. وبالمثل ، تشير β_2ر إلى التغير في متوسط الاستجابة لكل زيادة بمقدار الواحد في X_2 وذلك عندما يبقى X_1 ثابتا. ولرؤية هذا في مثالنا، لنفرض أن X_2 قد بقي عند المستوى $X_2 = 20$ فدالة الانحدار (7.3) هي الآن:

$$E\{Y\} = 20.0 + 0.95X_1 - 0.50(20) = (20.0\text{–}10.0) + 0.95X_1 \qquad (7.4)$$
$$= 10.0 + 0.95X_1$$

ونلاحظ أنه في حالة $X_2 = 20$ تكون دالة الاستجابة خطا مستقيما ميله 0.95. ويبقى الشيء نفسه صحيحا من أجل أي قيمة أخرى لـِ X_2 ؛ ومايختلف هو التقاطع مع سطح الاستجابة فقط. وبالتالي فإن $\beta_1 = 0.95$ تشير إلى زيادة متوسط الاستجابة بمقدار 0.95 عند زيادة في X_1 بمقدار الواحد، وذلك مع بقاء X_2 ثابتا، أيا كان مستوى X_2 وبكلام أقل دقة، نقول إن β_1 تشير إلى التغير في $E\{Y\}$ عند زيادة بمقدار الواحد في X_1 مع بقاء X_2 ثابتا.

وبصورة مماثلة، تشير $\beta_2 = -0.50$ في دالة الانحدار (7.3) إلى تناقص متوسط الاستجابة بمقدار 0.50 عند زيادة في X_2 بمقدار الواحد، وذلك مع بقاء X_1 ثابتا.

وعندما لايعتمد تأثير X_1 في متوسط الاستجابة على مستوى X_2، وفي المقابل لايعتمد تأثير X_2 على مستوى X_1 يُقال إن للمتغيرين المستقلين تأثيرات تجميعية أو إنهما لايتفاعلان. وهكذا فإن نموذج الانحدار من المرتبة الأولى (7.1) مصمم لمتغيرين مستقلين تأثيراهما على متوسط الاستجابة تجميعيان أو لايتفاعلان.

وكثيرا ماتدعى المعلمتان β_1 و β_2 معاملي انحدار جزئيين لأنهما يعكسان التأثير الجزئي لمتغير مستقل عندما يكون المتغير المستقل الآخر مشمولا في النموذج مع بقائه ثابتا.

مثال. لنفرض أن سطح الاستجابة في (7.3) يتعلق بمحطات خدمة شاملة حضرية لشركة نفط رئيسة. ويبين تأثيرات تنوع الخدمات وملاءمتها (X_1)، ومتوسط الزمن اللازم للوصول إلى عربة (X_2) على نسبة الجالونات المباعة فعلا من البنزين إلى المخزون الإجمالي من الجالونات (Y)، ويُعبر عن X_1 كرقم قياسي حيث المتوسط = 100، وعن X_2

بالثواني، وتُعرض Y كنسبة مئوية. وزيادة قياس ملاءمة الخدمات بنقطة واحدة مع بقاء متوسط زمن الوصول إلى عربة ثابتا، يؤدي إلى زيادة 0.95 بالمائة في النسبة المتوقعة للحالونات المباعة إلى المخزون الإجمالي. وإذا بقي قياس ملاءمة الخدمات ثابتا وزاد متوسط الزمن اللازم للوصول إلى عربة ثانية واحدة فإن النسبة المتوقعة للحالونات المباعة إلى المخزون الإجمالي تتناقص بمقدار نصف في المائة.

تعليقات

١- بالنسبة لنموذج انحدار سطحُ استجابته مستوٍ، يمكن استخدامه لذاته عندما يكون ذلك مناسبا، أو يمكن استخدامه كتقريب لسطح استجابة أكثر تعقيدا. ويمكن تقريب العديد من سطوح الاستجابة المعقدة بمستو، وذلك كتقريب جيد من أجل مَدَيين محدودين لِـ X_1 وX_2.

٢- يمكننا بسهولة إرساء معنى لِـ β_1 و β_2 باستخدام حساب التفاضل والتكامل، فبأخذ المشتقات الجزئية لسطح الاستجابة (7.2) بالنسبة لِـ X_1 ولِـ X_2 على التوالي نجد:

$$\frac{\partial E\{Y\}}{\partial X_1}=\beta_1 \qquad \frac{\partial E\{Y\}}{\partial X_2}=\beta_2$$

وتقيس المشتقات الجزئية معدل التغير في $E\{Y\}$ بالنسبة لأحد المتغيرين المستقلين عندما يبقى المتغير الآخر ثابتا.

نموذج من المرتبة الأولى بأكثر من متغيرين مستقلين

نعتبر الآن الحالة التي يوجد فيها $P - 1$ من المتغيرات المستقلة $X_1, \ldots, X_{p-1}$ ويدعى نموذج الانحدار:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \qquad (7.5)$$

نموذجا من المرتبة الأولى مع $P - 1$ متغيرا مستقلا. ويمكن أيضا كتابته على الشكل:

$$Y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad (7.5a)$$

أو يمكن كتابته، إذا جعلنا $X_{i0} \equiv 1$ ، على الشكل:

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad \text{حيث:} \quad X_{io} \equiv 1 \qquad (7.5b)$$

وبفرض $E\{\varepsilon_i\} = 0$، تكون دالة الاستجابة لنموذج الانحدار (7.5) هي:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} \quad (7.6)$$

ودالة الاستجابة هذه هي فوق مستوي، وهو مستوٍ بأكثر من بعدين. ولم يعُد من الممكن تصوير سطح الاستجابة هذا كما كنا قادرين على ذلك في الشكل (٧-١) في حالة متغيرين مستقلين. وتشير المعلمة β_k إلى التغير في متوسط الاستجابة $E\{Y\}$ عند زيادة بمقدار الواحد في المتغير المستقل X_k ، مع بقاء جميع المتغيرات المستقلة الأخرى في نموذج الانحدار ثابتة. ولنلاحظ أن تأثير أي متغير مستقل على متوسط الاستجابة في نموذج الانحدار (7.5) يبقى نفسه بصرف النظر عن المستويات التي ثبتنا عندها المتغيرات المستقلة الأخرى، وهكذا فإن نموذج الانحدار من المرتبة الأولى (7.5) مصمم لمتغيرات مستقلة تكون تأثيراتها على متوسط الاستجابة تجميعية وبالتالي فهي لاتتفاعل.

ملاحظة

إذا كان $P - 1 = 1$ يُختزل نموذج الانحدار في (7.5) إلى:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

وهو نموذج الانحدار الخطّي البسيط الذي درسناه في فصول سابقة.

نموذج الانحدار الخطّي العام

وبصورة عامة، ليس من الضروري أن تمثل المتغيرات X_1،...، X_{p-1} في نموذج الانحدار متغيرات مستقلة مختلفة، كما سنرى بعد قليل. ولذلك نعرّف نموذج الانحدار الخطّي العام ببساطة بدلالة متغيرات X، مع حدود خطأ طبيعية، كما يلي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (7.7)$$

حيث:

β_0، β_1،...، β_{p-1} هي معالم

X_{i1}،...، $X_{i,p-1}$ ثوابت معروفة

ε_i مستقلة $N(0,\sigma^2)$ ، $i = 1, \ldots, n$

وإذا جعلنا $X_{i0} \equiv 1$ فيمكن كتابة نموذج الانحدار (7.7) كما يلي:

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (7.7a)$$

حيث: $X_{i0} \equiv 1$

أو:

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad \text{حيث:} \quad X_{io} \equiv 1 \qquad (7.7b)$$

وباعتبار $E\{\varepsilon_i\} = 0$ فإن دالة الاستجابة لنموذج الانحدار (7.7) هي:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} \qquad (7.8)$$

وهكذا فإن نموذج الانحدار الخطي العام بحدود خطأ طبيعية، يتضمن أن المشاهدات Y_i هي متغيرات طبيعية مستقلة بمتوسط $E\{Y_i\}$ كما هو معطى في (7.8) وبتباين ثابت σ^2.

ويحيط هذا النموذج الخطّي العام بتشكيلة واسعة من الحالات، سنذكر قليلا منها الآن.

p - 1 متغيرا مستقلا. عندما تمثل المتغيرات X_1، X_2، ...، X_{p-1}، $p - 1$ من المتغيرات المستقلة المختلفة فإن نموذج الانحدار الخطي العام (7.7) هو، كما رأينا، نموذج من المرتبة الأولى لا يتضمن تأثيرات تفاعل بين المتغيرات المستقلة.

انحدار كثيرات الحدود. لنعتبر نموذج الانحدار المنحني بمتغير مستقل واحد:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i \qquad (7.9)$$

إذا فرضنا $X_{i1} = X_i$ و $X_{i2} = X_i^2$، فيمكن كتابة (7.9) كما يلي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

أي أن النموذج (7.9) هو حالة خاصة من نموذج الانحدار الخطّي العام (7.7). وبينما يوضح النموذج (7.9). نموذج انحدار منحن، دالة الاستجابة فيه تربيعية، فإن دوال استجابة على شكل كثيرات حدود من درجة أعلى هي أيضا حالات خاصة من نموذج انحدار خطي عام.

متغيرات محولة. لنعتبر النموذج:

$$\log Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad (7.10)$$

فسطح الاستجابة هنا معقّد، ومع ذلك يمكن التعامل مع النموذج (7.10) كنموذج انحدار خطّي عام. إذ لو جعلنا $Y_i' = \log Y_i$ فيمكن كتابة النموذج (7.10) كما يلي:

$$Y_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

وهو في صيغة نموذج الانحدار الخطّي العام (7.7) لقد اتفق أن كان المتغير التابع هو لوغاريتم Y.

ويمكن تحويل العديد من النماذج إلى نماذج انحدار خطّية عامة. وهكذا يمكن تحويل النموذج.

$$Y_i = \frac{1}{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i} \qquad (7.11)$$

إلى نموذج خطّي عام بجعل $Y_i'=1/Y_i$. ونجد عندئذ:

$$Y_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

تأثيرات تفاعل. لنعتبر نموذج الانحدار بمتغيرين مستقلين X_1 و X_2:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \qquad (7.12)$$

فمعنى β_1 و β_2 هنا يختلف عن معناهما المعطى سابقا بسبب وجود الحد الجدائي $\beta_3 X_{i1} X_{i2}$. ويمكن تبيان أن المتغير في متوسط الاستجابة المقابل لزيادة قدرها الواحد في X_1 مع بقاء X_2 ثابتا هو:

$$\beta_1 + \beta_3 X_2 \qquad (7.13)$$

وبصورة مماثلة، فإن التغير في متوسط الاستجابة المقابل لزيادة قدرها الواحد في X_2 مع بقاء X_1 ثابتا هو:

$$\beta_2 + \beta_3 X_1 \qquad (7.14)$$

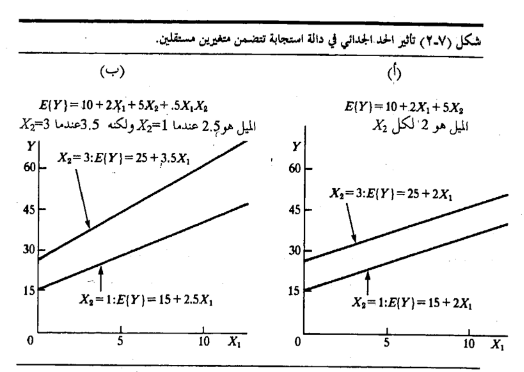
وبالتالي فإن كلا من تأثير X_1 من أجل مستوى معطى لِـ X_2 وتأثير X_2 من أجل مستوى معطى لِـ X_1 يعتمد، في نموذج الانحدار (7.12)، على مستوى المتغير المستقل الآخر.

ونوضّح في الشكل (٧-٢) تأثير الحد الجدائي في نموذج الانحدار (7.12). ففي الشكل (٧-٢)أ نعتبر دالة استجابة بدون حد جدائي:

$$E\{Y\} = 10 + 2X_1 + 5X_2$$

ونبين فيه دالة الاستجابة $E\{Y\}$ عندما يكون $X_2 = 1$ وعندما يكون $X_2 = 3$. ونلاحظ أن دالتي الاستجابة متوازيتان ـ أي أن متوسط الاستجابة يزداد بالمقدار نفسه وهو $\beta_1 = 2$ مع زيادة مقدارها الواحد في X_1 وذلك سواء أكان $X_2 = 1$ أو $X_2 = 3$.

شكل (٧-٢) تأثير الحد الجدائي في دالة استجابة تتضمن متغيرين مستقلين.



وفي الشكل (٧-٢)ب، نعتبر دالة الاستجابة نفسها ولكن بعد إضافة حد جدائي هو $0.5\,X_1X_2$:

$$E\{Y\} = 10 + 2X_1 + 5X_2 + 0.5X_1X_2$$

ونبين في (٧-٢)ب دالة الاستجابة $E\{Y\}$ عندما يكون $X_2 = 1$ و $X_2=3$، ونلاحظ، عند رسم ميلي دالتي الاستجابة في مقابل X_1، أنهما يختلفان الآن من أجل $X_2 = 1$ و$X_2=3$. فميل دالة الاستجابة عندما يكون $X_2 = 1$ هو من (7.13):

$$\beta_1 + \beta_3 X_2 = 2.+ 0.5(1) = 2.5$$

وعندما يكون $X_2 = 3$ فإن الميل يساوي:

$$\beta_1 + \beta_3 X_2 = 2 + 0.5(3) = 3.5$$

وهكذا فإن β_1 في نموذج الانحدار (7.12)، المتضمن لحد جدائي، لم يعد يشير إلى التغير في متوسط الاستجابة المقابل لزيادة مقدارها الواحد في X_1 من أجل أي مستو معطى لِ X_2 ، ففي هذا النموذج يعتمد ذلك التأثير على مستوى X_2 ونموذج الانحدار (7.12) المتضمن لحد جدائي هو إذا مصمم لمتغيرات مستقلة تتفاعل تأثيراتها على المتغير

التابع ويدعى الحد الجدائي $\beta_3 X_{i1} X_{i2}$ **حد التفاعل**. وبينما يبقى متوسط الاستجابة في نموذج الانحدار (7.12) دالة خطّية في X_1 عندما يكون X_2 ثابتا، إلا أن كلا من الجزء المقطوع لدالة الاستجابة وميلها يتغيران مع تغير القيمة التي ثبتنا عندها مستوى X_2. ويصح الشيء نفسه عند اعتبار متوسط الاستجابة كدالة في X_2 مع بقاء X_1 ثابتا.

وبالرغم من هذه التعقيدات في نموذج الانحدار (7.12) فلا يزال من الممكن اعتباره كنموذج انحدار خطّي عام. لتكن $X_{i3} = X_{i1} X_{i2}$ فيمكن أن نكتب (7.12) الآن كما يلي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

وهو في صيغة نموذج الانحدار الخطّي العام (7.7).

شكل (٧-٣) أمثلة إضافية لدوال استجابة

(أ) (ب)

ملاحظة

لاستنباط (7.13) و(7.14)، نشتق:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

بالنسبة لِـ X_1 وX_2 ، على الترتيب، فنجد:

$$\frac{\partial E\{Y\}}{\partial X_1}=\beta_1+\beta_3 X_2 \qquad \frac{\partial E\{Y\}}{\partial X_2}=\beta_2+\beta_3 X_1$$

مركب من الحالات. قد يضم نموذج انحدار عددا من العناصر التي ذكرناها آنفا، ونبقى قادرين مع ذلك على معالجته كنموذج انحدار خطّي عام. فلنعتبر نموذج انحدار بمتغيرين مستقلين كل منهما في صيغة تربيعية مع حد تفاعل:

$$Y_i=\beta_0+\beta_1 X_{i1}+\beta_2 X_{i1}^2+\beta_3 X_{i2}+\beta_4 X_{i2}^2+\beta_5 X_{i1} X_{i2}+\varepsilon_i \qquad (7.15)$$

ولنعرّف:

$$Z_{i1}=X_{i1} \quad Z_{i2}=X_{i1}^2 \quad Z_{i3}=X_{i2} \quad Z_{i4}=X_{i2}^2 \quad Z_{i5}=X_{i1} X_{i2}$$

فيمكن عندئذ كتابة نموذج الانحدار (7.15) كما يلي:

$$Y_i=\beta_0+\beta_1 Z_{i1}+\beta_2 Z_{i2}+\beta_3 Z_{i3}+\beta_4 Z_{i4}+\beta_5 Z_{i5}+\varepsilon_i$$

وهو في صيغة نموذج الانحدار الخطّي العام (7.7).

تعليقات

١- ينبغي أن يكون واضحا من الأمثلة المختلفة أن نموذج الانحدار الخطّي العام (7.7) غير مقصور على سطوح استجابة خطّية. ويشير مصطلح "**النموذج الخطّي**" إلى حقيقة أن (7.7) خطّي في المعالم ولايشير إلى شكل سطح الاستجابة.

٢- يوضّح الشكل (٧-٣) بعض سطوح الاستجابة المعقدة، عندما يكون لدينا متغيرين مستقلين بحيث يمكن تمثيلهما عن طريق نموذج الانحدار الخطّي العام (7.7).

التفاعلات وطبيعة سطح الاستجابة

قدمنا سابقا مفهوم تفاعل متغيرات مستقلة، وسنقدم الآن مزيدا من الإيضاح لكيفية اختلاف سطوح الاستجابة في حالة وجود تفاعل عنه في حالة عدم وجود تفاعل بين المتغيرات المستقلة.

ويتضمن الشكل (٧-٤)أ عرضا لسطح استجابة لايتفاعل فيه المتغيران المستقلان (متوسط درجة الحرارة الموسمي، ومقدار هطول المطر) على المتغير التابع (إنتاج الذرة). ويمكن رؤية غياب التفاعلات بالنظر إلى منحنيات إنتاج الذرة، من أجل متوسط موسمي معطى لدرجات الحرارة، كدالة في مقدار هطول المطر. فلهذه المنحنيات جميعا الشكل نفسه ولاتختلف عن بعضها إلا بمقدار ثابت. وهكذا فإن كل

إحداثي صادي على منحني إنتاج الذرة الموافق لمتوسط درجة حرارة 70° أعلى من الإحداثي الصادي المقابل له على منحني إنتاج الذرة الموافق لمتوسط درجة حرارة 78° بعدد ثابت من الوحدات.

وبصورة مكافئة يمكن ملاحظة غياب التفاعلات بالنظر إلى منحنيات إنتاج الذرة، من أجل مقدار معطى لهطول المطر، كدالة في درجة الحرارة. وثانية فإن لهذه المنحنيات الشكل نفسه ولاتختلف إلا بمقدار ثابت.

ويتضمن غياب التفاعلات، بالتالي، أنه يمكن التعبير عن متوسط الاستجابة $E\{Y\}$ على الشكل:

$$E\{Y\} = f_1(X_1) + f_2(X_2) \tag{7.16}$$

حيث يمكن أن تكون f_1 وf_2 أية دالتين، وليستا بالضرورة دالتين بسيطتين.

شكل (٧-٤) سطوح استجابة لمتغيرات مستقلة تجميعية ولمتغيرات مستقلة متفاعلة.

(أ) المتغيرات المستقلة غير متفاعلة

إنتاج الذرة كدالة في متوسط درجة الحرارة الموسمي ومقدار هطول المطر

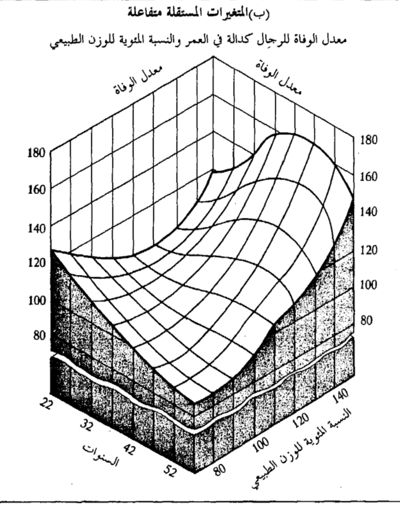
ويوضح الشكل (٧-٤)ب حالة يتفاعل فيها المتغيران المستقلان (العمر، النسبة المئوية للوزن العادي) على المتغير التابع (نسبة الوفيات). وهنا يختلف شكل منحني نسبة الوفيات كدالة في النسبة المئوية للوزن الطبيعي باختلاف الأعمار. فمن أجل

رجال أعمارهم 22 عاما نجد أن لكـل مـن الأشـخاص ذوي الـوزن المفـرط أو الـوزن المنخفض معدلات وفاة أعلى من المعدل العادي (العادي = 100) لذلك العمر. وعلـى الوجه الآخر، نجد أن معدل الوفاة لرجـال أعمـارهم 52 عامـا، هـو أعلـى مـن المعـدل العادي لذلك العمر بالنسبة للأشخاص ذوي الوزن المفرط وليس الأمر كذلك بالنسـبة للأشخاص ذوي الوزن المنخفض. وبصورة مماثلة فإن منحنيات معدلات الوفاة كدالـة في العمر تختلف في شكلها باختلاف الأوزان.

شكل ٧-٤ (تتمة)

(ب)المتغيرات المستقلة متفاعلة

معدل الوفاة للرجال كدالة في العمر والنسبة المئوية للوزن الطبيعي



المصدر: Reprinted, with prmission, from M. Ezekiel and K.A . Fox, *Methods of Correlation and Regression Analysis*, 3rd ed. (New York: John Wiley & Sons, 1959), pp. 349-50.

ويمكن إيضاح الفرق في شكل سطح الاستجابة بين حالتي تفاعل وعدم تفاعل المتغيرين المستقلين بطريقة أخرى أيضا ، ونقصد تمثيل سطح الاستجابة باستخدام مخططات منحنيات التساوي. وتبين مثل هذه المخططات، من أجل عدد من مستويات الاستجابة المختلفة، التراكيب المختلفة للمتغيرين المستقلين التي تُنتج مستوى الاستجابة نفسه. ويبين الشكل (٧-٥)أ مخطط منحنيات تساوي لسطح الاستجابة المصور في الشكل (٧-١):

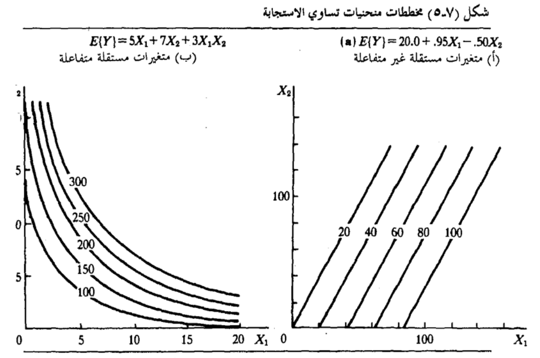
$$E\{Y\} = 20.0 + .95X_1 - .50X_2$$

ونلاحظ أن المتغيرين المستقلين لا يتفاعلان في دالة الاستجابة هذه وأن خطوط التساوي متوازية. ويبين الشكل (٧-٥)ب مخطط منحنيات التساوي لدالة الاستجابة:

$$E\{Y\} = 5X_1 + 7X_2 + 3X_1X_2$$

حيث يتفاعل المتغيران المستقلان ومنحنيات التساوي غير متوازية.

شكل (٧-٥) مخططات منحنيات تساوي الاستجابة



وبصورة عامة، تؤدي المتغيرات المستقلة غير المتفاعلة أو التجميعية إلى منحنيات تساوي متوازية، بينما تؤدي المتغيرات المستقلة المتفاعلة إلى منحنيات تساوي غير متوازية.

(٢-٧) نموذج انحدار خطّي عام بدلالة المصفوفات

سنقدم الآن النتائج الرئيسية لنموذج الانحدار الخطّي العام (7.7) بدلالة المصفوفات. وكما لاحظنا فإن هذا النموذج يحيط بتشكيلة واسعة من الحالات الخاصة، والنتائج التي سنقدمها قابلة للتطبيق على جميع هذه الحالات.

وإنها لخاصة رائعة من خواص جبر المصفوفات أن تبدو النتائج الخاصة بنموذج الانحدار الخطّي العام (7.7)، معبرا عنها بدلالة المصفوفات، مطابقة تماما لتلك الخاصة بنموذج الانحدار الخطّي البسيط (6.54). وماسيختلف هو فقط عدد درجات الحرية وثوابت أخرى تتصل بعدد المتغيرات المستقلة وبأبعاد بعض المصفوفات. وبالتالي سنكون قادرين على تقديم النتائج بصورة مختصرة جدا.

ومن المؤكد أن رمز المصفوفة يمكن أن يُخفي تعقيدات حسابية هائلة فمعكوس مصفوفة **A** أبعادها 10×10 يتطلب مقادير هائلة من الحساب، ومع ذلك فهي تقدم ببساطة على الشكل $\mathbf{A}^{-1}$. وسبب تأكيدنا على جبر المصفوفات هو أنه يشير إلى الخطوات الذهنية الجوهرية من خطوات الحل. وستتم الحسابات الفعلية جميعها، باستثناء الحالات الأكثر بساطة، باستخدام حاسب يدوي مبرمج أو حاسب آلي. وبالتالي فإنه لايعنينا ما إذا كان $(\mathbf{X'X})^{-1}$ يمثل إيجاد معكوس مصفوفة أبعادها 2×2 أو 10×10. والنقطة المهمة هي معرفة ماذا يمثل معكوس مصفوفة.

وللتعبير عن نموذج الانحدار الخطّي العام (7.7):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

بدلالة المصفوفات نحتاج إلى تعريف المصفوفات التالية:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} \qquad (7.17a)$$

$$\underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2,p-1} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{bmatrix} \qquad (7.17b)$$

(7.17c) (7.17d) (7.17)

$$\underset{p\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_{p-1} \end{bmatrix} \qquad \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

لاحظ أن **Y** و ε هما المتجهان نفسهما كما في الانحدار الخطّي البسيط. ويتضمن المتجه β معالم انحدار إضافية وتتضمن المصفوفة **X** عمودا من الأعداد 1 بالإضافة إلى عمود من القيم الـ n لكل من المتغيرات X في نموذج الانحدار وعددها $P - 1$. ودليل الصف لكل عنصر X_{ik} في المصفوفة **X** يحدد التكرار أو المشاهدة، ويحدد دليل العمود المتغير X.

وبدلالة المصفوفات يصبح نموذج الانحدار الخطّي العام (7.7):

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \tag{7.18}$$

حيث:

Y متجه الاستجابات

β متجه المعالم

X مصفوفة من الثوابت

ε متجه من المتغيرات العشوائية الطبيعية بتوقع $E\{\varepsilon\} = \mathbf{0}$ ومصفوفة تغاير $\sigma^2\{\varepsilon\} = \sigma^2\mathbf{I}$

وبالتالي، فإن للمتجه العشوائي **Y** توقعا:

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \mathbf{X}\boldsymbol{\beta} \tag{7.18a}$$

ومصفوفة تغاير **Y** هي:

$$\underset{n\times n}{\sigma^2\{\mathbf{Y}\}} = \sigma^2\mathbf{I} \tag{7.18b}$$

(٧-٣) مقدرات المربعات الدنيا

لنرمز بـ **b** لمتجه معاملات الانحدار المقدّرة b_0، b_1، ...، b_{p-1}:

$$\underset{p\times 1}{\mathbf{b}}=\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ . \\ . \\ . \\ b_{p-1} \end{bmatrix} \tag{7.19}$$

فمعادلات المربعات الدنيا الناظمية لنموذج الانحدار الخطّي العام (7.18) هي:

$$\mathbf{X'Xb}=\mathbf{X'Y} \tag{7.20}$$

ومقدِّرات المربعات الدنيا هي:

$$\underset{p\times 1}{\mathbf{b}}=\underset{p\times p}{(\mathbf{X'X})^{-1}}\underset{p\times 1}{\mathbf{X'Y}} \tag{7.21}$$

وفي حالة نموذج الانحدار (7.18) تكون مقدِّرات المربعات الدنيا هذه مقدِّرات الإمكانية العظمى أيضا ولها جميع الخواص المذكورة في الفصل الثاني: فهي غير منحازة، وغير منحازة بتباين أصغري، ومتّسقة، وكافية.

(٧-٤) القيم التوفيقية والرواسب

لنرمز بـ $\hat{\mathbf{Y}}$ لمتجه القيم التوفيقية $\hat{Y}_i$ وبـ **e** لمتجه حدود الراسب $e_i=Y_i-\hat{Y}_i$:

$$\underset{n\times 1}{\hat{\mathbf{Y}}}=\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ . \\ . \\ . \\ \hat{Y}_n \end{bmatrix} \quad (7.22b) \qquad \underset{n\times 1}{\mathbf{e}}=\begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix} \quad (7.22a) \tag{7.22}$$

فيمكن تمثيل القيم التوفيقية على الشكل:

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \mathbf{Xb} \tag{7.23}$$

وحدود الراسب على الشكل:

$$\underset{n\times 1}{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{Xb} \tag{7.24}$$

ويمكن التعبير عن متجه القيم التوفيقية $\hat{\mathbf{Y}}$ بدلالة مصفوفة القبّعة **H** كما يلي:

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \mathbf{HY} \tag{7.25}$$

حيث:

$$\underset{n\times n}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{7.25a}$$

وبصورة مماثلة، يمكن التعبير عن متجه الرواسب كما يلي:

$$\underset{n\times 1}{\mathbf{e}} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \tag{7.26}$$

ومصفوفة تغاير تباين الرواسب هي:

$$\underset{n\times n}{\sigma^2\{\mathbf{e}\}} = \sigma^2(\mathbf{I}-\mathbf{H}) \tag{7.27}$$

وهي تُقدَّر بما يلي:

$$\underset{n\times n}{s^2\{\mathbf{e}\}} = MSE\,(\mathbf{I}-\mathbf{H}) \tag{7.28}$$

(٧ـ٥) نتائج تحليل التباين

مجموع مربعات ومتوسط مربعات

ومجاميع المربعات في تحليل التباين بدلالة المصفوفات هي:

$$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{JY} = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \tag{7.29}$$

$$SSE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb}) = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} \tag{7.30}$$
$$= \mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}$$

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{JY} = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \tag{7.31}$$

حيث **J** هو $n\times n$ مصفوفة من المقادير 1، كنا عرفناها في (6.18) و**H** مصفوفة القبعة كما عرفناها في (7.25a).

$SSTO$، كالمعتاد، لها $n - 1$ درجة من الحرية تترافق معهـا. وSSE لها $n - P$ مـن درجات الحرية تترافق معها، باعتبار أننا نحتاج إلى تقدير p من المعالم في دالـة الانحـدار للنموذج (7.18). وأخيرا SSR لها $P - 1$ درجـة مـن الحريـة تـترافق معهـا، ممثلـة لعـدد المتغيرات X وهي X_1، . . .، X_{p-1}.

ويبين الجدول (٧-١) هذه النتائج لتحليل التباين بالإضافة إلى متوسطي المربعات MSR وMSE:

$$MSR = \frac{SSR}{p-1} \tag{7.32}$$

جدول (٧-١) جدول تباين لنموذج الانحدار الخطّي العام (7.18)

مصدر التغير	SS	df	MS
الانحدار	$SSR = \mathbf{b'X'Y} - \left(\frac{1}{n}\right)\mathbf{Y'JY}$	$p - 1$	$MSR = \frac{SSR}{p-1}$
الخطأ	$SSE = \mathbf{Y'Y} - \mathbf{b'X'Y}$	$n - p$	$MSE = \frac{SSE}{n-p}$
المجموع	$SSTO = \mathbf{Y'Y} - \left(\frac{1}{n}\right)\mathbf{Y'JY}$	$n - 1$	

$$MSE = \frac{SSE}{n-p} \tag{7.33}$$

وتوقع MSE هو σ^2، كما في حالة الانحدار الخطّي البسيط. وتوقع MSR هـو σ^2 مضافا إليه كمية غير سالبة. وعلى سبيل المثال، عندما $P - 1 = 2$، يكون لدينا:

$$E\{MSR\} = \sigma^2 + \left[\begin{array}{l}\beta_1^2 \Sigma(X_{i1} - \overline{X}_1)^2 + \beta_2^2 \Sigma(X_{i2} - \overline{X}_2)^2 \\ \qquad + 2\beta_1\beta_2 \Sigma(X_{i1} - \overline{X}_1)(X_{i2} - \overline{X}_2)\end{array}\right] \Big/ 2$$

لاحظ أنه إذا كان كل من β_1 وβ_2 مساويا للصفر فإن $E\{MSR\} = \sigma^2$ وفيما عدا ذلك يكون: $E\{MSR\} > \sigma^2$.

الاختبار F لعلاقة انحدار

ولاختبار ما إذا كانت توجد علاقة انحدار بين المتغير التابع Y ومجموعة المتغـيرات

X ، وهي X_1،...، X_{p-1}، أي للاختيار بين البديلين:

$$H_0: \beta_1 = \beta_2 = ... = \beta_{p-1} = 0 \tag{7.34a}$$

H_a : ليست كل β_k ($k=1,...,p-1$) تساوي الصفر

نستخدم إحصاءة الاختبار:

$$F^* = \frac{MSR}{MSE} \tag{7.34b}$$

وقاعدة القرار عند ضبط الخطأ من النوع الأول عند α، هي:

$$\text{إذا كان } F^* \leq F(1-\alpha; p-1, n-p) \text{ استنتج } H_0 \tag{7.34c}$$
$$\text{إذا كان } F^* > F(1-\alpha; p-1, n-p) \text{ استنتج } H_a$$

ووجود علاقة انحدار لذاتها لا يؤكد بالطبع إمكانية الوصول إلى تنبؤات مفيدة باستخدام هذه العلاقة.

ونلاحظ أنه عندما يكون p-1=1 فإن الاختبار يُختزل إلى الاختبار F في (3.61) الخاص باختبار ما إذا كان $\beta_1 = 0$ في انحدار خطّي بسيط.

معامل التحديد المتعدّد

يُعرّف معامل التحديد المتعدد، ونرمز له بـ R^2، كما يلي:

$$\mathbf{R}^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \tag{7.35}$$

وهو يقيس التخفيض النسبي في التغير الكلي لِ Y الذي يترافق مع استخدام مجموعة المتغيرات X وهي X_1،...،X_{p-1} ويُختزل معامل التحديد المتعدد R^2 إلى معامل التحديد r^2 في (3.71) الخاص بانحدار خطّي بسيط، عندما يكون p-1 = 1، أي عندما يوجد متغير مستقل واحد في نموذج الانحدار (7.18) ولدينا، تماما كما في حالة r^2:

$$0 \leq R^2 \leq 1 \tag{7.36}$$

ويفترض R^2 القيمة 0 عندما تكون جميع المقادير $b_k = 0$ $(k = 1,..., p-1)$ مساوية للصفر. ويأخذ R^2 القيمة 1 عندما تقع جميع المشاهدات Y على سطح الاستجابة التوفيقي مباشرة، أي عندما يكون $Y_i = \hat{Y}_i$ من أجل جميع قيم i.

تعليقات

١- للتمييز بين معاملي التحديد في حالتي انحدار بسيط وانحدار متعدّد، سندعو r^2، من الآن فصاعدا، معامل التحديد البسيط.

٢- يمكن تبيان إمكانية النظر إلى معامل التحديد المتعدّد R^2 كمعامل تحديد بسيط r^2 بين الاستجابة Y_i والقيم التوفيقية $\hat{Y}_i$.

٣- القيم الكبيرة لـ R^2 لاتتضمن بالضرورة أن النموذج الذي تمّ توفيقه هو نموذج مفيد. وعلى سبيل المثال، يمكن أن تكون المشاهدات قد أُخذت عند مستويات قليلة فقط للمتغيرات المستقلة. وبالرغم من ارتفاع R^2 في هذه الحالة فقد لايكون النموذج مفيدا لأن معظم التنبؤات ستحتاج إلى توسع في الاستقراء خارج منطقة المشاهدات. وأيضا ، حتى عندما يكون R^2 كبيرا، فقد يكون MSE كبيرا إلى حد لاتكون معه الاستقراءات مفيدة إذا أردنا لدقة هذه الاستقراءات أن تكون دقة عالية.

٤- إضافة المزيد من المتغيرات المستقلة إلى النموذج يمكن أن يؤدي فقط إلى زيادة R^2، ولايخفضها أبدا، لأن SSE لا يمكن أن تصبح أبدا أكبر مع مزيد من المتغيرات المستقلة، ولأن $SSTO$ تبقى دائما نفسها من أجل مجموعة معطاة من الاستجابات. وبما أنه يمكننا، في الغالب، جعل R^2 كبيرة باعتماد عدد كبير من المتغيرات المستقلة، فيُقترح أحيانا استخدام مقياس معدَّل يأخذ في الاعتبار عدد المتغيرات المستقلة في النموذج. ومعامل التحديد المتعدد المعدل، ويرمز له بـ R_a^2 ، يعدِّل R^2 بتقسيم كل مجموع مربعات على عدد درجاته من الحرية؛ وهكذا نجد:

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO} \tag{7.37}$$

ويمكن أن يصبح معامل التحديد المتعدد هذا أصغر عند إدخال متغير مستقل آخر إلى النموذج. لأن النقص في SSE يمكن أن يكون أكثر من أن يعوض عن نقص درجة حرية في المقام $n - P$.

معامل الارتباط المتعدّد

معامل الارتباط المتعدّد R هو الجذر التربيعي الموجب لِـ R^2

$$R=\sqrt{R^2} \tag{7.38}$$

وهو يساوي في القيمة المطلقة معامل الارتباط r في (3.73) لارتباط بسيط عندما يكون $P - 1 = 1$ ، أي عندما يوجد متغير مستقل واحد في نموذج الانحدار (7.18).

ملاحظة

من الآن فصاعدا، سندعو r معامل الارتباط البسيط لتمييزه عن معامل الارتباط المتعدّد.

(٦-٧) استدلالات حول معالم الانحدار

مقدرات المربعات الدنيا في $\mathbf{b}$ غير منحازة:

$$E\{b\} = \beta \tag{7.39}$$

ومصفوفة التغاير $\sigma^2\{\mathbf{b}\}$

$$\underset{p\times p}{\sigma^2\{\mathbf{b}\}} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0,b_1\} & \dots & \sigma\{b_0,b_{p-1}\} \\ \sigma\{b_1,b_0\} & \sigma^2\{b_1\} & \dots & \sigma\{b_1,b_{p-1}\} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sigma\{b_{p-1},b_0\} & \sigma\{b_{p-1},b_1\} & \dots & \sigma^2\{b_{p-1}\} \end{bmatrix} \tag{7.40}$$

معطاة بالعلاقة:

$$\underset{p\times p}{\sigma^2\{\mathbf{b}\}} = \sigma^2 (\mathbf{X'X})^{-1} \tag{7.41}$$

ومصفوفة التغاير المقدَّرة $\mathbf{s}^2\{\mathbf{b}\}$:

$$\underset{p\times p}{s^2\{\mathbf{b}\}} = \begin{bmatrix} s^2\{b_0\} & s\{b_0,b_1\} & \dots & s\{b_0,b_{p-1}\} \\ s\{b_1,b_0\} & s^2\{b_1\} & \dots & s\{b_1,b_{p-1}\} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ s\{b_{p-1},b_0\} & s\{b_{p-1},b_1\} & \dots & s^2\{b_{p-1}\} \end{bmatrix} \tag{7.42}$$

معطاة بالعلاقة:

$$\underset{p\times p}{\mathbf{s}^2\{\mathbf{b}\}} = MSE\,(\mathbf{X'X})^{-1} \tag{7.43}$$

ويمكن أن نحصل من $s^2\{\mathbf{b}\}$ على $s^2\{b_0\}$ و $s^2\{b_1\}$، أو أي تبـاين آخـر نحتاجـه، أو أيـة تغايرات نحتاجها.

التقدير بفترة لِـ β_k

في نموذج الانحدار (7.18) ذي الخطأ الطبيعي، لدينا:

$$\frac{b_k-\beta_k}{s\{b_k\}} \sim t(n-p) \qquad k = 0,1,\ldots, p-1 \tag{7.44}$$

وبالتالي فإن حدي الثقة لِـ β_k بمعامل ثقة $1-\alpha$ هما:

$$b_k \pm t(1-\alpha/2;\, n-p)s\{b_k\} \tag{7.45}$$

اختبارات β_k

تجري اختبارات β_k بالطريقة المعتادة. فلاختبار:

$$H_0\colon \beta_k = 0$$
$$H_a\colon \beta_k \neq 0 \tag{7.46a}$$

يمكن استخدام إحصاءة الاختبار:

$$t^* = \frac{b_k}{s\{b_k\}} \tag{7.46b}$$

وقاعدة القرار هي:

إذا كان $|t^*| \le t(1-\alpha/2\,;\ n-P)$ استنتج H_0

فيما عدا ذلك استنتج H_a. (7.46c)

ويمكن الحصول على قوة الاختبار t كما شرحنا في الفصـل الثـالث، مـع تعديـل عدد درجات الحرية ليصبح $n-P$.

وكما في حالة الانحدار الخطّي البسيط، يمكن أيضا القيام باختبار ما إذا كان $\beta_k = 0$ أ لا في نماذج الانحدار المتعدّد باستخدام الاختبار F. ونناقش هذا الاختبار في الفصل الثامن.

استدلالات مشتركة

يمكـن اسـتخدام فـترات الثقـة المشـتركة لبونفرّونـي لتقديـر عـدد مـن معـاملات الانحدار في آن واحد. وإذا أردنا تقدير g من المعالم بصورة مشـتركة (حيـث $g \le P$

فإن حدي الثقة بمعامل ثقة عائلي α - 1 هما:

$$b_k \pm Bs\{b_k\} \tag{7.47}$$

حيث:

$$B = t(1 - \alpha/2g; n - p) \tag{7.47a}$$

ونناقش في الفصل الثامن اختبارات تتعلق بمجموعة جزئية من معالم الانحدار.

(٧-٧) استدلالات حول متوسط الاستجابة

التقدير بفترة لِـ $E\{\mathbf{Y}_h\}$

من أجل قيم معطاة لِـ X_1،...،X_{p-1}، ولنرمز لها بـ X_{h1}،...،$X_{h,p-1}$ نرمز لمتوسط الاستجابة بـ $E\{Y_h\}$. ونعرّف المتجه $\mathbf{X}_h$:

$$\underset{p\times 1}{\mathbf{X}_h} = \begin{bmatrix} 1 \\ X_{h1} \\ X_{h2} \\ . \\ . \\ . \\ X_{h,p-1} \end{bmatrix} \tag{7.48}$$

فيكون متوسط الاستجابة المراد تقديره هو:

$$E\{Y_h\} = \mathbf{X}_h' \boldsymbol{\beta} \tag{7.49}$$

ومتوسط الاستجابة المقدَّر الموافق لِـ $\mathbf{X}_h$، ونرمز له بـ $\hat{Y}_h$ هو:

$$\hat{Y}_h = \mathbf{X}_h' \mathbf{b} \tag{7.50}$$

وهذا المقدِّر غير منحاز:

$$E\{\hat{\mathbf{Y}}_h\} = \mathbf{X}_h' \boldsymbol{\beta} = E\{Y_h\} \tag{7.51}$$

وتباينه هو:

$$\sigma^2\{\hat{\mathbf{Y}}_h\} = \sigma^2 \mathbf{X}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_h = \mathbf{X}_h' \sigma^2 \{\mathbf{b}\} \mathbf{X}_h \tag{7.52}$$

لاحظ أن التباين $\sigma^2\{\hat{Y}_h\}$ هو دالة في تباينات معاملات الانحدار $\sigma^2\{b_k\}$ وفي التغايرات $\sigma\{b_k, b_{k'}\}$ لأزواج من معاملات الانحدار، وذلك تماما كما في الانحدار الخطّي البسيط. والتباين المقدر $s^2\{\hat{Y}_h\}$ معطى بالعلاقة:

$$s^2\{\hat{\mathbf{Y}}_h\} = MSE\,(\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h) = X_h' s^2\{\mathbf{b}\}\mathbf{X}_h \qquad (7.53)$$

وحدا الثقة لـِ $E\{Y_h\}$ بمعامل ثقة $1 - \alpha$ هما:

$$\hat{\mathbf{Y}}_h \pm t(1-\alpha/2; n-p)s\{\hat{\mathbf{Y}}_h\} \qquad (7.54)$$

فترات ثقة متزامنة لعدة متوسطات استجابة

عندما نرغب في تقدير عدة متوسطات استجابة $E\{Y_h\}$ مقابلة لمتجهات مختلفة $\mathbf{X}_h$ يمكن استخدام أسلوبين أساسيين للتحكم في معامل الثقة العائلي $1 - \alpha$:

١- استخدم حدي ثقة من نوع (*Working - Hotelling*) وورکنج ـ هوتلّنج للمتجهات $\mathbf{X}_h$ المعنية:

$$\hat{Y}_h \pm Ws\{\hat{Y}_h\} \qquad (7.55)$$

حيث:

$$W^2 = pF(1 - \alpha; p, n - p) \qquad (7.55a)$$

٢- استخدم فترات الثقة المتزامنة لبونفرّوني. وعندما يُراد القيام بـ g تقدير بفترة، فإن حدّي الثقة لبونفرّوني هما:

$$\hat{Y}_h \pm Bs\{\hat{Y}_h\} \qquad (7.56)$$

حيث:

$$B = t(1 - \alpha/2g; n - p) \qquad (7.56a)$$

ومن أجل أي تطبيق بعينه، ينبغي مقارنة W و B لرؤية أيهما يؤدي إلى فترات ثقة أقصر. وإذا لم تكن مستويات $\mathbf{X}_h$ محددة سلفا، ولكنها تتحدد مع مضي التحليل، فمن الأفضل استخدام الحدين من نوع وورکنج - هوتلّنج في (7.55).

اختبار F حول نقص التوفيق

لاختبار ما إذا كانت دالة استجابة الانحدار المتعدد:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1}X_{p-1}$$

تمثل سطح استجابة مناسب لمجموعة من البيانات فإننا نحتاج، كما في تحليل الانحدار

الخطّي البسيط، إلى مشاهدات مكررة. والمشاهدات المكررة في الانحدار المتعدّد هي مشاهدات متكررة لقيمة Y المقابلة لمستويات كل من المتغيرات X التي تبقى ثابتة من تكرار إلى آخر. وهكذا فإن المشاهدات المكررة، في حالة متغيرين مستقلين، يتطلب بقاء كل من X_1 و X_2 عند مستو ثابت من مشاهدة لقيمة Y إلى مشاهدة أخرى.

والإجراءات التي وصفناها في الفصل الرابع والمتعلقة باختبار F حول نقص التوفيق هي اجراءات قابلة للتطبيق في الانحدار المتعدد. وحالما نحصل على جدول التحاين، المبين في الجدول (٧-١)، نحلل SSE إلى مركبتي خطأ بحت ونقص توفيق. ونحصل على مجموع مربعات الخطأ البحت $SSPE$ بأن نحسب أولا، ولكل زمرة من المشاهدات المكررة، مجموع مربعات انحرافات المشاهدات Y عن متوسط الزمرة، حيث تبقى قيم المتغيرات X نفسها في كل من زمر التكرارات. فلنفرض وجود c من زمر التكرارات بمجموعات متميزة من مستويات المتغيرات X ، ولنرمز لمتوسط المشاهدات Y في الزمرة j بـ $\overline{Y}_j$. فمجموع المربعات للزمرة j معطى في (4.12)، ومجموع مربعات الخطأ البحت هو مجموع هذه المجاميع من المربعات كما أُعطي في (4.11). ومجموع مربعات نقصْ التوفيق $SSLF$ يساوي الفرق $SSE - SSPE$ كما أشرنا في (4.19).

وعدد درجات الحرية المرافق لِ $SSPE$ هو $n - c$ ، وعدد درجات الحرية المرافق لِ $SSLF$ هو $(n - P) - (n - c) = c - P$.

ويجري الاختبار F كما وصفنا في الفصل الرابع، ولكن بدرجات من الحرية معدلة عن تلك المعروضة هناك. ولاختبار البدائل:

$$\begin{aligned} &H_0: E\{Y\} = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1} \\ &H_a: E\{Y\} \neq \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1} \end{aligned} \tag{7.57a}$$

تكون إحصاءة الاختبار المناسبة:

$$F^* = \frac{SSLF}{c-p} \div \frac{SSPE}{n-c} = \frac{MSLF}{MSPE} \tag{7.57b}$$

حيث $SSLF$ و$SSPE$ معطيان في (4.19) و(4.11)، على الترتيب، وقاعدة القرار المناسبة هي:

$$\begin{aligned} &\text{إذا كانت } F^* \leq F(1-\alpha; c-p, n-c) \text{ استنتج } H_0 \\ &\text{إذا كانت } F^* > F(1-\alpha; c-p, n-c) \text{ استنتج } H_a \end{aligned} \tag{7.57c}$$

(٧-٨) تنبؤات بمشاهدات جديدة

تنبؤ بمشاهدة جديدة $Y_{h(new)}$

حدا التنبؤ بمعامل ثقة α - 1 لمشاهدة $Y_{h(new)}$ جديدة مقابلة لـ $\mathbf{X}_h$ متجه القيم المحددة للمتغيرات X ، هما:

$$\hat{Y}_h \pm t(1-\alpha/2; n-p)\, s\{Y_{h(new)}\} \tag{7.58}$$

حيث

$$\begin{aligned} s^2\{Y_{h(new)}\} &= MSE + s^2\{\hat{Y}_h\} = MSE + X'_h s^2\{\mathbf{b}\}\mathbf{X}_h \\ &= MSE\,(1 + X'_h\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h) \end{aligned} \tag{7.58a}$$

تنبؤ بمتوسط m من المشاهدات الجديدة عند $\mathbf{X}_h$

عندما نرغب في اختيار m مشاهدة جديدة عند المستوى $\mathbf{X}_h$ للمتغيرات X ونريد التنبؤ بمتوسطها $\overline{Y}_{h(new)}$ فإن حدي التنبؤ بمعامل ثقة α - 1 هما:

$$\hat{Y}_h \pm t(1-\alpha/2; n-p)\, s\{\overline{Y}_{h(new)}\} \tag{7.59}$$

حيث:

$$\begin{aligned} s^2\left\{\overline{Y}_{h(new)}\right\} &= \frac{MSE}{m} + s^2\{\hat{Y}_h\} = \frac{MSE}{m} + X'_h s^2\{\mathbf{b}\}\mathbf{X}_h \\ &= MSE\left(\frac{1}{m} + X'_h\,(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right) \end{aligned} \tag{7.59a}$$

تنبؤ بـ g من المشاهدات الجديدة

تعطى حدود شفّيه المتزامنة للتنبؤ بـ g من المشاهدات الجديدة عند g من المستويات المختلفة لـ $\mathbf{X}_h$ بمعامل ثقة عائلي α - 1 بالعبارة:

$$\hat{Y}_h \pm S s\{\mathbf{Y}_{h(new)}\} \tag{7.60}$$

حيث:

$$S^2 = gF(1-\alpha; g, n-p) \tag{7.60a}$$

و $s^2\{\mathbf{Y}_{h(new)}\}$ معطى بالعلاقة (7.58a).

وبصورة بديلة، يمكن استخدام حدود بونفرّوني المتزامنة للتنبؤ. وهي، من أجل g من التنبؤات وبمعامل ثقة عائلي α - 1، معطاة بالعبارة:

$$\hat{Y}_h \pm Bs\{Y_{h(\text{new})}\} \quad (7.61)$$

حيث:

$$B = t\,(1-\alpha/2g;\, n-p) \quad (7.61a)$$

ومقارنة S و B سلفا لأي استخدام بعينه ستشير إلى أي الطريقتين ستؤدي إلى فترات أضيق للتنبؤ.

(٧-٩) رسومات الرواسب، تشخيصات أخرى، وتدابير علاجية

الطرق التشخيصية التي ناقشناها في الفصل الرابع للانحدار الخطي البسيط، هي أيضا طرق مفيدة للانحدار الخطّي المتعدد. وهكذا فإن رسومات الصناديق، ورسومات الزمن، ورسومات الجذع والورقة، والرسومات النقطية لكل من المتغيرات المستقلة، يمكن أن تقدم معلومات مساعدة وتمهيدية حول هذه المتغيرات. وبصورة مماثلة، فإن رسومات الانتشار للمتغير التابع في مقابل كل من المتغيرات المستقلة، يمكن أن تعين في تحديد طبيعة وقوة العلاقة بين المتغير المستقل والمتغير التابع، وفي التعرف على ثغرات في نقاط تمثيل البيانات بالإضافة إلى التعرف على نقاط البيانات القاصية. ورسومات الانتشار لكل متغير مستقل مقابل كل من المتغيرات المستقلة الأخرى هي رسومات معينة في دراسة العلاقات بين المتغيرات المستقلة، وفي إيجاد ثغرات، وتحرّي المشاهدات القاصية.

ورسم الرواسب في مقابل القيم التوفيقية مفيد لتثمين صلاحية دالة الانحدار، وثبات تباين حدود الخطأ، بالإضافة إلى تقديم معلومات عن المشاهدات القاصية. وبصورة مماثلة، يمكن أن يقدم رسم الرواسب في مقابل الزمن معلومات تشخيصية حول ارتباطات ممكنة بين حدود الخطأ. وتفيد رسومات *Box* (بوكس)، ورسومات الاحتمال الطبيعي للرواسب، في تفحّص ما إذا كانت حدود الخطأ تتوزع بصورة معقولة وفق التوزيع الطبيعي.

وبالإضافة إلى ذلك، ينبغي رسم الرواسب في مقابل كل من المتغيرات المستقلة. ويمكن أن يقدّم كل من هذه الرسومات معلومات إضافية حول صلاحية دالة الانحدار بالنسبة لذلك المتغير المستقل (مثلا، ما إذا كنا نحتاج تمثيلا منحنيا لتأثير ذلك المتغير)، وحول تغيرات ممكنة في مقدار تباين الخطأ فيما يتعلق بذلك المتغير المستقل.

وأخيرا، ينبغي رسم الرواسب في مقابل متغيرات مستقلة مهمة حُذفت من النموذج لرؤية ما إذا كان للمتغيرات المحذوفة تأثيرات إضافية مهمة على المتغير التابع لم نتعرف عليها بعد من خلال نموذج الانحدار. وينبغي أيضا رسم الرواسب في مقابل حدود التفاعل غير المشمولة في نموذج الانحدار، مثل X_1X_2 ، X_1X_3 وX_2X_3 ، لرؤية ما إذا كنا نحتاج، في النموذج، لبعض حدود التفاعل هذه أو لها جميعا.

والتدابير العلاجية الموصوفة في الفصل الرابع هي أيضا تدابير قابلة للتطبيق في الانحدار المتعدد. وإذا تطلب الأمر نموذجا أكثر تعقيدا يسلّم بوجود تأثيرات منحنية أو تأثيرات تفاعل، فيمكن توسيع نموذج الانحدار المتعدد ليشمل هذه التأثيرات. وعلى سبيل المثال يمكن إضافة X_2^2 كمتغير ليأخذ في الاعتبار تأثيرا منحنيا لِ X_2، أو يمكن إضافة X_1X_3 كمتغير اعترافا بوجود تأثير تفاعل بين X_1 و X_3 على المتغير التابع. وبصورة بديلة، يمكن القيام بتحويلات على المتغير التابع و/أو المتغيرات المستقلة، متّبعين في ذلك المبادىء التي ناقشناها في الفصل الرابع، لعلاج أية عيوب في النموذج. ويمكن أن تكون التحويلات في المتغير التابع مُعِيْنة عندما تكون توزيعات حدود الخطأ ملتوية بوضوح، وتباين حدود الخطأ غير ثابت. كما يمكن أن تكون التحويلات في بعض المتغيرات المستقلة مُعِيْنة عندما تكون تأثيرات هذه المتغيرات تأثيرات منحنية. وبالإضافة إلى ذلك، يمكن أن تكون التحويلات على **Y** و/أو المتغيرات المستقلة مُعِيْنة في حذف تأثيرات التفاعل أو تخفيضها تخفيضا هائلا.

وكما في الانحدار الخطّي البسيط، نحتاج إلى الاطمئنان إلى فائدة التحويلات باستخدام رسومات الرواسب وأدوات التشخيص الأخرى، وذلك لتحديد ما إذا كان نموذج الانحدار المتعدد مناسبا للبيانات بعد التحويل.

ملاحظة

وصفنا في الفصل الرابع أسلوب بوكس - كوكس لتحديد تحويل قوة مناسب لـِ Y، في نماذج الانحدار البسيط. وهذا الأسلوب قابل للتطبيق أيضا على نماذج الانحدار المتعدد. وبالاضافة إلى ذلك طوّر بوكس وتيدول (*Box & Tidwell*) المرجع [7.1] أسلوبا تكراريا للتعرف على تحويلات القوى المناسبة لكل من المتغيرات المستقلة في نموذج انحدار متعدد.

(٧-١٠) مثال - انحدار متعدد مع متغيرين مستقلين

في هذه الفقرة سنوضح بالتفصيل تطبيقا للانحدار المتعدد مع متغيرين مستقلين. وسنوضح عددا من الإجراءات التشخيصية المختلفة، وعدة أنواع من الاستقراءات التي يمكن القيام بها في هذا التطبيق.

جدول (٧-٢) بيانات أساسية - مثال شركة زارثان

المنطقة i	المبيعات بالكروز من المرطبات (الكروز = ١٢ دزينة) Y_i	المجتمع الهدف (بآلاف الأشخاص) X_{i1}	الدخل الفردي المصرّح به (بالدولارات) X_{i2}
1	162	274	2,450
2	120	180	3,254
3	223	375	3,802
4	131	205	2,838
5	67	86	2,347
6	169	265	3,782
7	81	98	3,008
8	192	330	2,450
9	116	195	2,137
10	55	53	2,560
11	252	430	4,020
12	232	372	4,427
13	144	236	2,660
14	103	157	2,088
15	212	370	2,605

إطار المسألة

تبيع شركة زارثان "كريم" خاص بالجلد وذلك حصرا في محلات "أدوات الزينة". وتشمل أعمالها خمس عشرة منطقة تسويقية، وتهتم بالتنبؤ بمبيعات المنطقة. ويتضمن الجدول (٧-٢) بيانات تتعلق بالمبيعات في كل منطقة، بالإضافة إلى بيانات عن كل منطقة حول المجتمع الهدف والدخل الفردي المصرح به. وسنعالج المبيعات كمتغير تابع Y، والمجتمع الهدف والدخل الفردي المصرح به كمتغيرين مستقلين X_1

وX_2 على الترتيب، في استكشاف لإمكانية التنبؤ بمبيعات المنطقة من المجتمع الهدف والدخل الفردي المصرح به. ويُتوقع أن يكون نموذج الانحدار من المرتبة الأولى:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad (7.62)$$

مع حدود خطأ طبيعية، نموذجا مناسبا.

جدول (٧-٣) المصفوفتان Y و X - مثال شركة زارثان

$$\mathbf{Y} = \begin{bmatrix} 162 \\ 120 \\ 223 \\ 131 \\ 67 \\ 169 \\ 81 \\ 192 \\ 116 \\ 55 \\ 252 \\ 232 \\ 144 \\ 103 \\ 212 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 274 & 2{,}450 \\ 1 & 180 & 3{,}254 \\ 1 & 375 & 3{,}802 \\ 1 & 205 & 2{,}838 \\ 1 & 86 & 2{,}347 \\ 1 & 265 & 3{,}782 \\ 1 & 98 & 3{,}008 \\ 1 & 330 & 2{,}450 \\ 1 & 195 & 2{,}137 \\ 1 & 53 & 2{,}560 \\ 1 & 430 & 4{,}020 \\ 1 & 372 & 4{,}427 \\ 1 & 236 & 2{,}660 \\ 1 & 157 & 2{,}088 \\ 1 & 370 & 2{,}605 \end{bmatrix}$$

حسابات أساسية

يبين الجدول (٧-٣) المصفوفتين Y وX لمثال شركة زارثان التوضيحي. وسنحتاج إلى:

-١

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & .. & 1 \\ 274 & 180 & \ldots & 370 \\ 2{,}450 & 3{,}254 & \ldots & 2{,}605 \end{bmatrix} \begin{bmatrix} 1 & 274 & 2{,}450 \\ 1 & 180 & 3{,}254 \\ . & . & . \\ & . & . \\ 1 & 370 & 2{,}605 \end{bmatrix}$$

وهذا يعطي:

$$\mathbf{X'X} = \begin{bmatrix} 15 & 3,626 & 44,428 \\ 3,626 & 1,067,614 & 11,419,181 \\ 44,428 & 11,419,181 & 139,063,428 \end{bmatrix} \quad (7.63)$$

٢-

$$\mathbf{X'Y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 274 & 180 & \dots & 370 \\ 2,450 & 3,254 & \dots & 2,605 \end{bmatrix} \begin{bmatrix} 162 \\ 120 \\ \cdot \\ \cdot \\ \cdot \\ 212 \end{bmatrix}$$

وهذا يعطي:

$$\mathbf{X'Y} = \begin{bmatrix} 2,259 \\ 647,107 \\ 7,096,619 \end{bmatrix} \quad (7.64)$$

٣-

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 15 & 3.626 & 44,428 \\ 3,626 & 1,067,614 & 11,419,181 \\ 44,428 & 11,419,181 & 139,063,428 \end{bmatrix}^{-1}$$

وباستخدام (6.23)، نعرّف

$$\begin{array}{lll} a = 15 & b = 3,626 & c = 44,428 \\ d = 3,626 & e = 1,067,614 & f = 11,419,181 \\ g = 44,428 & h = 11,419,181 & k = 139,063,428 \end{array}$$

بحيث يكون:

$$\begin{aligned} Z &= 14,497,044,060,000 \\ A &= 1.246348416 \\ B &= 0.0002129664176 \end{aligned}$$

وهلم جرا. ونحصل على:

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 1.2463484 & 2.1296642E-4 & -4.1567125E-4 \\ 2.1296642E-4 & 7.7329030E-6 & -7.0302518E-7 \\ -4.1567125E-4 & -7.0302518E-7 & 1.9771851E-7 \end{bmatrix} \quad (7.65)$$

ونلاحظ أن بعض النتائج في المصفوفة $(\mathbf{X'X})^{-1}$ معطاة في هيئة أسية (E format) حيث ترمز E -4، مثلا، لِ $10^{-4}=1/10^{4}$ وهكذا ترمز 2.1296642 E - 4 لِ 0.00021296642.

المكافيء الجبري. لاحظ أن **X'X** لنموذج الانحدار من المرتبة الأولى (7.62) بمتغيرين مستقلين هو:

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & X_{n1} & X_{n2} \end{bmatrix}$$

أو

$$\mathbf{X'X} = \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i2} & \sum X_{i2}X_{i1} & \sum X_{i2}^2 \end{bmatrix} \tag{7.66}$$

وهكذا نجد في مثال شركة زارثان:

$$n = 15$$
$$\sum X_{i1} = 274 + 180 + \dots = 3,626$$
$$\sum X_{i1}X_{i2} = 274(2,450) + 180(3,254) + \dots = 11,419,181$$

وقد حُسبت هذه العناصر في (7.63).

لاحظ أيضا أن **X'Y** لنموذج الانحدار من المرتبة الأولى (7.62) بمتغيرين مستقلين هو:

$$\mathbf{X'Y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{i1}Y_i \\ \sum X_{i2}Y_i \end{bmatrix} \tag{7.67}$$

ولدينا، في مثال شركة زارثان:

$$\sum \mathbf{Y}_i = 162 + 120 + \dots = 2,259$$
$$\sum X_{i1}\mathbf{Y}_i = 274(162) + 180(120) + \dots = 647,107$$

$$\sum X_{i2}Y_i = 2{,}450(162) + 3{,}254(120) + \ldots = 7{,}096{,}619$$

وهذه العناصر هي العناصر التي حُسبت في (7.64).

دالة الانحدار المقدَّرة

ونحصل بسهولة على تقديرات المربعات الدنيا **b** باستخدام (7.21)، وبالاستفادة من النتائج الأساسية التي حصلنا عليها في (7.64) و(7.65):

$$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

$$= \begin{bmatrix} 1.2463484 & 2.1296642E-4 & -4.1567125E-4 \\ 2.1296642E-4 & 7.7329030E-6 & -7.0302518E-7 \\ -4.1567125E-4 & -7.0302518E-7 & 1.9771851E-7 \end{bmatrix} \times \begin{bmatrix} 2{,}259 \\ 647{,}107 \\ 7{,}096{,}619 \end{bmatrix}$$

$$= \begin{bmatrix} 3.4526127900 \\ 0.4960049761 \\ 0.009199080867 \end{bmatrix}$$

وهكذا:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 3.4526127900 \\ 0.4960049761 \\ 0.009199080867 \end{bmatrix}$$

وتكون دالة الانحدار المقدَّرة:

$$\hat{Y} = 3.453 + 0.96X_1 + 0.00920X_2$$

وتشير دالة الانحدار المقدَّرة هذه إلى أنه يتوقع زيادة متوسط مبيعات العبوات (الكروزات) بمقدار 0.496 كروز (جملة عبوات أو كروز = ١٢ دزينة) وذلك عندما يزيد المجتمع الهدف بمقدار ألف مع بقاء الدخل الفردي المصرح به ثابتا، ويتوقع زيادة متوسط مبيعات العبوات بمقدار 0.0092 كروز عندما يزيد الدخل الفردي المصرح به دولارا واحدا ، مع بقاء المجتمع على حاله.

الشكل الجبري للمعادلات الناظمية. ومن (7.66) و(7.67) يمكن الحصول بسهولة على المعادلات الناظمية في حالة متغيرين مستقلين، في شكلها الجبري فلدينا:

$$(\mathbf{X'X})\mathbf{b} = \mathbf{X'Y}$$

$$=\begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i2} & \sum X_{i2}X_{i1} & \sum X_{i2}^2 \end{bmatrix}\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}=\begin{bmatrix} \sum Y_i \\ \sum X_{i1}Y_i \\ \sum X_{i2}Y_i \end{bmatrix}$$

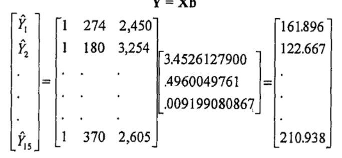
ومنها نحصل على المعادلات الناظمية:

$$\begin{aligned} \sum Y_i &= nb_0 + b_1 \sum X_{i1} + b_2 \sum X_{i2} \\ \sum X_{i1}Y_i &= b_0 \sum X_{i1} + b_1 \sum X_{i1}^2 + b_2 \sum X_{i1}X_{i2} \\ \sum X_{i2}Y_i &= b_0 \sum X_{i2} + b_1 \sum X_{i1}X_{i2} + b_2 \sum X_{i2}^2 \end{aligned} \qquad (7.68)$$

الرواسب والقيم التوفيقية

ولفحص مصداقية نموذج الانحدار (7.62) للبيانات التي بين أيدينا، نحتاج إلى القيم التوفيقية $\hat{Y}_i$ والرواسب $e_i = Y_i - \hat{Y}_i$ ونحصل من (7.23) على:

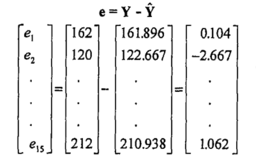
$$\hat{\mathbf{Y}} = \mathbf{Xb}$$



$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{Y}_{15} \end{bmatrix} = \begin{bmatrix} 1 & 274 & 2,450 \\ 1 & 180 & 3,254 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & 370 & 2,605 \end{bmatrix} \begin{bmatrix} 3.4526127900 \\ .4960049761 \\ .009199080867 \end{bmatrix} = \begin{bmatrix} 161.896 \\ 122.667 \\ \cdot \\ \cdot \\ \cdot \\ 210.938 \end{bmatrix}$$

وبالإضافة إلى ذلك، نجد من (7.24):

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$$



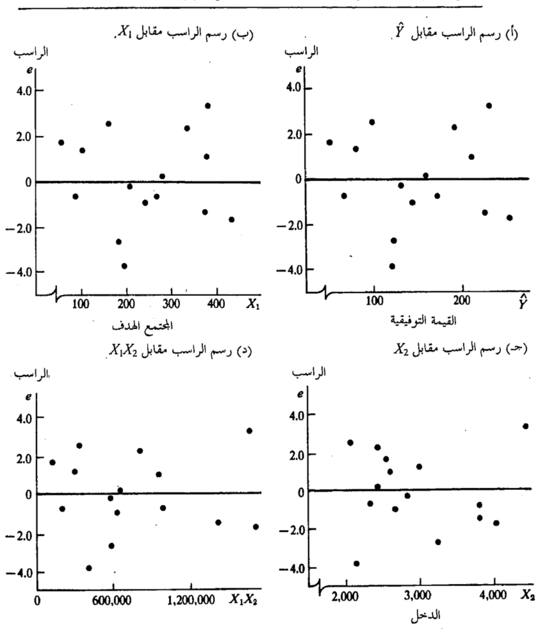
$$\begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_{15} \end{bmatrix} = \begin{bmatrix} 162 \\ 120 \\ \cdot \\ \cdot \\ \cdot \\ 212 \end{bmatrix} - \begin{bmatrix} 161.896 \\ 122.667 \\ \cdot \\ \cdot \\ \cdot \\ 210.938 \end{bmatrix} = \begin{bmatrix} 0.104 \\ -2.667 \\ \cdot \\ \cdot \\ \cdot \\ 1.062 \end{bmatrix}$$

تحليل مصداقية نموذج

نبدأ تحليلنا لصلاحية نموذج الانحدار (7.62) لبيانات مثال شركة زارثان بتأمل رسم الرواسب e في مقابل القيم التوفيقية $\hat{Y}$ في الشكل (٧-٦)أ. ولايقترح هذا الرسم أية انحرافات نمطية عن مستوي الاستجابة، أو أية تغيرات في تباين حدود الخطأ

مع تغير مستوى $\hat{Y}$ ورسومات الرواسب e في مقابل X_1 وX_2 في الشكلين (٧-٦)ب و(٧-٦)جـ ، على الترتيب، متسقة تماما مع نتائج جودة التوفيق لدالة الاستجابة ومع نتائج ثبات تباين حدود الخطأ.

شكل (٧-٦) رسومات الرواسب التشخيصية ـ مثال شركة زارثان

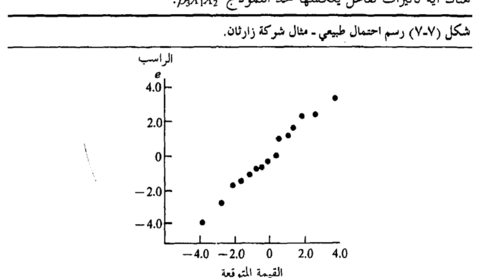


وكثيرا ماتوجد، في تطبيقات الانحدار المتعدد، إمكانية حضور لتأثيرات تفاعل. ولفحص هذه الإمكانية في مثال شركة زارثان، رسمنا الرواسب e في مقابل حد التفاعل X_1X_2 في الشكل (٧-٦)د. ووجود نمط نظامي في هذا الرسم كان سيقترح إمكانية وجود تفاعل، وبحيث تكون دالة استجابة من النوع:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

أكثر ملاءمة عندئذ. ولكن الشكل (٧-٦)د لايقدم أي نمط نظامي وبالتالي لايبدو أن هناك أية تأثيرات تفاعل يعكسها حد النموذج $\beta_3 X_1 X_2$.

شكل (٧-٧) رسم احتمال طبيعي ـ مثال شركة زارثان.



وأخيرا ، يتضمن الشكل (٧-٧) رسم احتمال طبيعي للرواسب، ويبدو الرسم خطّيا إلى درجة مقبولة مما يتفق مع توزيع طبيعي لحدود الخطأ. ومعامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت فرض التوزيع الطبيعي هو 0.993 وتساعد هذه القيمة المرتفعة (انظر الجدول ٤-٣) في التأكيد على معقولية الاستنتاج بأن حدود الخطأ تتبع، بصورة مقبولة، التوزيع الطبيعي.

وبما أن البيانات في مثال شركة زارثان لاتنطوي على أي تتابع زمني فالرسومات الزمنية غير واردة هنا، وهكذا تدعم جميع التشخيصات استخدام نموذج الانحدار (7.62) لبيانات شركة زارثان.

تحليل تباين

لاختبار ما إذا كانت المبيعات تتعلق بالمجتمع وبالدخل الفردي المصرح به، ننشيء في الجدول (٧-٤) جدول تحاين. والكميات الرئيسة التي نحتاجها هي:

$$\mathbf{Y'Y} = [162 \quad 120 \ldots 212]\begin{bmatrix}162\\120\\ \cdot \\ \cdot \\ \cdot \\212\end{bmatrix}$$

جدول (٧-٤) جدول تحاين - مثال شركة زارثان

مصدر التغير	*SS*	*df*	*MS*
الانحدار	$SSR = 53{,}844.716$	2	$MSR = 26{,}922.358$
الخطأ	$SSE = 56.884$	12	$MSE = 4.740$
المجموع	$SSTO = 53{,}901.600$	14	

$$= (162)^2 + (120)^2 + \ldots + (212)^2$$
$$= 394{,}107.000$$

$$\left(\frac{1}{n}\right)\mathbf{Y'JY} = \frac{1}{15}[162 \quad 120 \ldots 212]\begin{bmatrix}1 & 1 & 1 & \ldots & 1\\1 & 1 & 1 & \ldots & 1\\ \cdot & \cdot & \cdot & & \cdot\\ \cdot & \cdot & \cdot & & \cdot\\ \cdot & \cdot & \cdot & & \cdot\\1 & 1 & 1 & \ldots & 1\end{bmatrix}\begin{bmatrix}162\\120\\ \cdot \\ \cdot \\ \cdot \\212\end{bmatrix}$$

$$= \frac{(2{,}259)^2}{15} = 340{,}205.400$$

وهكذا نجد:

$$SSTO = \mathbf{Y'Y} - \left(\frac{1}{n}\right)\mathbf{Y'JY} = 394{,}107.000 - 340{,}205.400 = 53{,}901.600$$

وباستخدام نتائجنا في (7.64) نجد:

$$SSE = \mathbf{Y'Y} - \mathbf{b'X'Y}$$

$$= 394{,}107.000 - [3.4526127900 \;\; .4960049761 \;\; .009199080867] \times \begin{bmatrix} 2{,}59 \\ 647{,}107 \\ 7{,}096{,}619 \end{bmatrix}$$

$$= 394{,}107.000 - 394{,}050.116 = 56.884$$

وأخيرا نحصل بالطرح على:

$$SSR = SSTO - SSE = 53{,}901.600 - 56.884 = 53{,}844.716$$

وقد أدخلت متوسطات المربعات وأعداد درجات الحرية في الجدول (٧-٤). لاحظ أنه وجب علينا تقدير ثلاث معالم، وبالتالي يترافق مع SSE عدد من درجات الحرية يساوي 12 = 3 - 15. وأيضا عدد درجات الحرية المرافقة لِ SSR هو 2 أي عدد المتغيرات X في النموذج.

اختبار علاقة الحدار. ولاختبار ما إذا كانت المبيعات تتعلق بالمجتمع وبالدخل الفردي المصرح به، أي اختبار:

$$H_0 : \beta_1 = 0 \quad , \quad \beta_2 = 0$$

H_a: ليس كل من β_2 و β_1 مساوٍ للصفر

نستخدم إحصاءة الاختبار ($7.34b$):

$$F^* = \frac{MSR}{MSE} = \frac{26{,}922.358}{4.740} = 5{,}680$$

وبفرض $\alpha = 0.05$، نحتاج إلى $F(.95; 2{,}12) = 3.89$. وبما أن $F^* = 5680 > 3.89$ نستنتج H_a، أي أن للمبيعات صلة بالمجتمع وبالدخل الفردي المصرح به. والقيمة -P لهذا الاختبار هي أقل من 0.001، إذ نلاحظ من الجدول (أ-٤) أن $F(.999;2{,}12)=13.0$.

ويبقى علينا رؤية ما إذا كانت علاقة الانحدار مفيدة للقيام بتنبؤات عن المبيعات أو للوصول إلى تقديرات لمتوسط المبيعات.

معامل التحديد المتعدد. في مثالنا، لدينا من (7.53):

$$R^2 = \frac{SSR}{SSTO} = \frac{53{,}844.716}{53{,}901.600} = 0.9989$$

وهكذا عند اعتبار المتغيرين المستقلين: المجتمع والدخـل الفـردي المصـرح بـه، ينخفـض التغير في المبيعات 99.9 بالمائة.

عبارة جبرية لِـ *SSE*. مجموع مربعات الخطأ، في حالة متغـيرين مسـتقلين، في صورتـه الجبرية هو:

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} = \sum Y_i^2 - \begin{bmatrix} b_0 & b_1 & b_2 \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_{i1} Y_i \\ \sum X_{i2} Y_i \end{bmatrix}$$

أو:

$$SSE = \sum Y_i^2 - b_0 \sum Y_i - b_1 \sum X_{i1} Y_i - b_2 \sum X_{i2} Y_i \qquad (7.69)$$

لاحظ كيف تشكّل هذه العبارة تعميما مباشرا للعبارة (2.24*a*) في حالة متغير مستقل واحد.

تقدير معالم الانحدار

لا تثير المعلمـة β_0 اهتمـام شـركة زارثـان باعتبارهـا واقعـة بعيـدا خـارج مجـال النموذج. ونرغب في تقدير β_1 و β_2 معا بمعامل ثقة عائلي 0.90. وسنستخدم حـدي الثقة المتزامنين، لبونفرّوني المذكورين في (7.47).

ونحتاج أولا إلى تقدير مصفوفة التغاير $\mathbf{s}^2\{\mathbf{b}\}$:

$$\mathbf{s}^2\{\mathbf{b}\} = MSE\,(\mathbf{X}'\mathbf{X})^{-1}$$

و *MSE* معطى في الجدول (٧-٤)، وقد حصلنا في (7.65) على $(\mathbf{X}'\mathbf{X})^{-1}$ وبالتالي:

$$\mathbf{s}^2\{\mathbf{b}\} = 4.7403 \begin{bmatrix} 1.2463484 & 2.1296642E-4 & -4.1567125E-4 \\ 2.1296642E-4 & 7.7329030E-6 & -7.0302518E-7 \\ -4.1567125E-4 & -7.0302518E-7 & 1.9771851E-7 \end{bmatrix}$$

$$= \begin{bmatrix} 5.9081 & 1.0095E-3 & -1.9704E-3 \\ 1.0095E-3 & 3.6656E-5 & -3.3326E-6 \\ -1.9704E-3 & -3.3326E-6 & 9.3725E-7 \end{bmatrix}$$

والتباينان المقدَّران اللذان نحتاجهما هما:

$s^2\{b_1\} = .000036656$ أو $s\{b_1\} = .006054$

$s^2\{b_2\} = .00000093725$ أو $s\{b_2\} = .0009681$

وبعد ذلك، نحتاج، في حالة $g = 2$، أي تقديرين في آن واحد، إلى:

$$B = t[1 - .10 / 2(2);\ 12] = t(.975;12) = 2.179$$

وهكذا يكون حدا الثقة المتزامنان هما (0.006054) 2.179 ± 0.4960

و(0.0009681) 2.179 + 0.009199 وهذا ينتج فترتي الثقة:

$$0.483 \le \beta_1 \le 0.509$$
$$0.0071 \le \beta_2 \le 0.0113$$

ونستنتج بمعامل ثقة عائلي 0.90 أن تقع β_1 بين 0.483 و 0.509 وأن تقع β_2 بين 0.0071 و0.0113.

لاحظ أن فترتي الثقة المتزامنين تقترحان أن كلا من β_1 و β_2 موجبة، مما يتفق مع التوقعات النظرية بأن المبيعات ينبغي أن تزيد مع عدد أكبر من السكان ومع دخل فردي أعلى، عند إبقاء المتغير الآخر ثابتا.

تقدير متوسط الاستجابة

ترغب شركة زارثان في تقدير (متوسط) المبيعات المتوقعة في منطقة عدد سكانها X_{h1} يساوي 220 ألف نسمة، والدخل الفردي المصرح به X_{h2} يساوي 2500 دولارا.

لنعرّف:

$$X = \begin{bmatrix} 1 \\ 220 \\ 2{,}500 \end{bmatrix}$$

فالتقدير النقطي لمتوسط المبيعات هو من (7.50):

$$\hat{Y}_h = X_h' \mathbf{b} = \begin{bmatrix} 1 & 220 & 2{,}500 \end{bmatrix} \begin{bmatrix} 3.4526 \\ 0.4960 \\ 0.009199 \end{bmatrix} = 135.57$$

وباستخدام (7. 53) والنتائج في (7.70) نجد التباين المقدَّر:

$$s^2\{\hat{Y}_h\} = X_h' s^2\{\mathbf{b}\} X_h$$
$$= [1 \quad 220 \quad 2{,}2500]$$
$$\times \begin{bmatrix} 5.9081 & 1.0095E-3 & -1.9704E-3 \\ 1.0095E-3 & 3.6656E-5 & -3.3326E-6 \\ -1.9704E-3 & -3.3326E-6 & 9.3725E-7 \end{bmatrix} \begin{bmatrix} 1 \\ 220 \\ 2{,}500 \end{bmatrix} = 0.46638$$

أو:

$$s\{\hat{\mathbf{Y}}_h\} = 0.68292$$

وبفرض أن معامل الثقة للتقدير بفترة لـ $E\{Y_h\}$ هو 0.95، نحتاج إلى $t(0.975;12)=2.179$، ومن (7.54) نحصل على حدي الثقة $135.57\pm2.179(0.68292)$ وفترة الثقة لـ$E\{Y_h\}$ هي إذا:

$$134.1 \le E\{Y_h\} \le 137.1$$

وهكذا، وبمعامل ثقة 0.95، نقدر أن متوسط المبيعات من الكروزات في منطقة عدد سكانها 220 ألفا ودخلها الفردي المصرح به 2500 دولارا يتراوح بين 134.1 كروزا و137.1 كروزا من العبوات.

النسخة الجبرية للتباين المقدّر $s^2\{\hat{Y}_h\}$. بما أن (7.53) تعطينا:

$$s^2\{\hat{Y}_h\} = X_h' s^2\{b\} X_h$$

فنستنتج، في حالة متغيرين مستقلين، أن:

$$s^2\{\hat{Y}_h\} = s^2\{b_0\} + X_{h1}^2 s^2\{b_1\} + X_{h2}^2 s^2\{b_2\} + 2X_{h1}s\{b_0,b_1\} + 2X_{h2}s\{b_0,b_2\} + 2X_{h1}X_{h2}s\{b_1,b_2\} \quad (7.71)$$

وعندما نعوض في (7.71) مستخدمين التباينات والتغايرات المقدّرة كما وردت في (7.70) نحصل على النتيجة أعلاه، ونقصد $s^2\{\hat{Y}_h\}=0.46638$.

حدا التنبؤ لمشاهدات جديدة

ترغب شركة زارثان في التنبؤ بالمبيعات في منطقتين لهما المعطيات التالية:

	منطقة A	منطقة B
X_{h1}	220	375
X_{h2}	2,500	3,500

ولتحديد أي فترتي التنبؤ المتزامنتين أفضل هنا، سنحسب S كما هي معطاة في (7.60a) وB كما هي معطاة في (7.61a)، وذلك في حالة $g = 2$، وبافتراض 0.90 كمعامل ثقة عائلي لنجد:

$$S^2 = 2F(.90; 2, 12) = 2(2.81) = 5.62 \qquad S = 2.37$$

و :

$$B = t[1 - 0.10/2(2); 12] = t(0.975; 12) = 2.179$$

وبالتالي يكون حدا بونفرّوني أكثر كفاءة هنا.

وللمنطقة A، سنستخدم النتائج التي وجدناها عند تقدير متوسط المبيعات، باعتبار أن مستويي المتغيرين المستقلين هما نفساهما هنا. ولدينا سابقا:

$$\hat{Y}_h = 135.57 \qquad s^2\{\hat{Y}_h\} = 0.46638 \qquad MSE = 4.7403$$

وبالتالي لدينا من (7.58a):

$$s^2\{Y_{h(new)}\} = \boldsymbol{MSE} + s^2\{\hat{Y}_h\} = 4.7403 + 0.46638 = 5.20668$$

أو:

$$s\{Y_{h(new)}\} = 2.28182$$

وبطريقة مماثلة لدينا في حالة المنطقة B (الحسابات غير مبينة):

$$\hat{Y}_h = 221.65 \qquad s\{Y_{h(new)}\} = 2.34536$$

ووجدنا سابقا أن معامل بونفرّوني هو $B = 2.179$. ومن (7.61) نجد بالتالي أن حدود التنبؤ المتزامنة لبونفرّوني بمعامل ثقة 0.90 هي:

$$135.57 \pm 2.179(2.28182) \text{ و } 221.65 \pm 2.179(2.34536)$$

مما يؤدي إلى فترتي التنبؤ المتزامنتين:

منطقة A : $130.6 \le \mathbf{Y}_{h(new)} \le 140.5$

منطقة B : $216.5 \le \mathbf{Y}_{h(new)} \le 226.8$

ونتنبأ، بمعامل ثقة عائلي 0.90، أن المبيعات في هاتين المنطقتين ستكون ضمن الحدود المشار إليها. وتعتبر شركة زارثان أن حدود التنبؤ هذه دقيقة بصورة كافية، وبالتالي فهي مفيدة.

مُخرجات الحاسب

يتضمن الشكل (٧-٨) مخرجات حاسب توضيحية لمثال شركة زارثان، وقد تمّ الحصول عليها باستخدام برنامج *GLM* (*General Linear Model*) من حزمة الحاسب *SAS* (*Statistical Analysis System*). (مرجع [7.2]) وتختلف مخرجات تحليل الانحدار من حيث هيئتها من برنامج حاسب إلى آخر، ويمكن رؤية ذلك بمقارنة المُخْرَج في الشكل (٧-٨) بمخرجات أخرى مقدّمة في فصول سابقة. وعلى أي حال فإن المعلومات الأساسية المقدمة في المخرجات المختلفة تبقى، بصورة رئيسة، نفسها في حزم الانحدار الإحصائية الرئيسة.

شكل (٧-٨) مخرجات الحاسب في مثال شركة زارثان (*SAS* ، مرجع [7.2])

THE X'X MATRIX

	INTERCEPT	TARGTP	INCOME
INCERCEPT ← X_0	15.00	3626.00	44428.00
TARGTP ← X_1	3626.00	1067614.00	11419181.00
INCOME ← X_2	44428.00	11419181.00	139063428.00

X'X INVERSE MATRIX

	INTERCEPT	TARGTP	INCOME
INTERCEPT	1.24634842	0.00021297	-0.00041567
TARGTP	0.00021297	0.00000775	-0.00000070
INCOME	-0.00041567	-0.00000070	0.00000020

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	3.45261279	1.42	0.1809	2.43065049
TARGTP	0.49600498	81.92	0.0001	0.00605444
INCOME	0.00919908	9.50	0.0001	0.00096811
	↑ b_k	↑ $t_k^* = b_k/s\{b_k\}$	↑ Two-sided P-value	↑ $s\{b_k\}$

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	2	SSR → 53844.71643444	MSR → 26922.35821722	F^* → 5679.47
ERROR	12	SSE → 56.88354556	MSE → 4.74029713	
CORRECTED TOTAL	14	53901.60000000 ↑ $SSTO$		

One-sided P-value

PR > F	R-SQUARE
→ 0.0001	0.998945 ← R^2

STD DEV

2.17722234 ← $\sqrt{MSE}$

OBSERVATION	Y_i OBSERVED VALUE	$\hat{Y}_i$ PREDICTED VALUE	e_i RESIDUAL
1	162.00000000	161.89572437	0.10427563
2	120.00000000	122.66731763	-2.66731763
3	223.00000000	224.42938429	-1.42938429
4	131.00000000	131.24062439	-0.24062439
5	67.00000000	67.69928353	-0.69928353
6	169.00000000	169.68485530	-0.68485530
7	81.00000000	79.73193570	1.26806430
8	192.00000000	189.67200303	2.32799697
9	116.00000000	119.83201895	-3.83201895
10	55.00000000	53.29052354	1.70947646
11	252.00000000	253.71505760	-1.71505760
12	232.00000000	228.69079490	3.30920510
13	144.00000000	144.97934226	-0.97934226
14	103.00000000	100.53307489	2.46692511
15	212.00000000	210.93805961	1.06194039

وقد كتبنا حواشي على صفحة المخرج في الشكل (٧-٨) لبيان الصلة مع الرموز المستخدمة في هذا الكتاب. ويتضمن القطاعان الأولان من المعلومات نتائج تحليل الانحدار الوسيطة في شكل مصفوفي، وعلى وجه الخصوص المصفوفتين $\mathbf{X'X}$ و$(\mathbf{X'X})^{-1}$. والعنوان *Intercept* في هاتين المصفوفتين يشير إلى $X_{i0} \equiv 1$ في نموذج الانحدار البديل (7.7*b*).

ويقدم القطاع التالي معلومات عن معاملات الانحدار المقدَّرة b_k. ويبين، على التوالي، التقديرات b_k، إحصاءات الاختبارات $t_k^* = b_k / s\{b_k\}$ لاختبار ما إذا كان $\beta_k = 0$ أم لا، القيم -*P* ذات الجانبين لإحصاءات الاختبارات، والانحرافات المعيارية المقدَّرة $s\{b_k\}$.

ويتضمن القطاع الرابع معلومات التحاين: جدول التحاين، القيمة F^* لاختبار ما إذا كانت توجد علاقة انحدار أم لا، القيمة - *P* لهذا الاختبار، $\sqrt{MSE}$ و R^2. ويبين القطاع الأخير القيم الملحوظة Y_i، القيم التوفيقية $\hat{Y}_i$، والرواسب e_i.

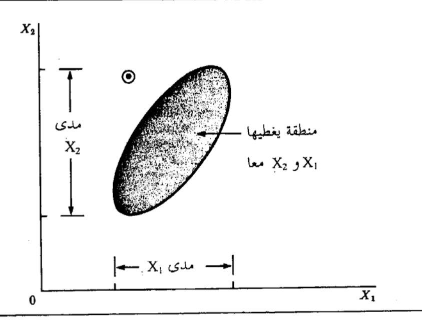
وبسبب تدوير الأرقام العشرية لاتتطابق بعض النتائج في الشكل (٧-٨) تماما مع النتائج المقابلة المعطاة سابقا. وفي هذا المضمار تنبغي ملاحظة أن حزم انحدار حاسوبية مختلفة يمكن أن تؤدي إلى نتائج مختلفة، إلى حد ما، وذلك بسبب أن النتائج النهائية يجري تقريبها بدرجات مختلفة من الدقة، ولسبب، ربما كان أكثر أهمية، و يتمثل في أن تدوير الأرقام العشرية لايتم بالدرجة نفسها من الجودة في جميع الحزم. وعلى وجه الخصوص، عندما يوجد عدد من المتغيرات المستقلة، ويرتبط بعض منها ارتباطا عاليا، فإن تدوير الأرقام العشرية يمكن أن يشكّل مصدرا خطرا من مصادر الصعوبة. وإنها لسياسة حكيمة أن نتحرى حزمة انحدار حاسوبية قبل استخدامها، بمقارنة مخرجاتها في مسألة اختبار، مثلا، مع نتائج نعرف أنها دقيقة.

تحذير من استقراءات خفية خارج مجال النموذج

وقبل اختتام هذا التوضيح لتحليل انحدار متعدد، ينبغي أن نحذر ثانية من القيام بتقديرات أو تنبؤات خارج مجال النموذج. والخطورة تأتي، بالطبع، من أن النموذج

قد لايكون مناسبا عند تعميمه إلى خارج منطقة المشاهدات، وفي الانحدار المتعدد، يسهل، على وجه الخصوص أن نتيه في هذه المنطقة، باعتبار أنها معرّفة بصورة مشتركة بمستويات X_1، ...، X_{p-1} وهكذا لايمكن الاكتفاء بالنظر إلى مدى كل من المتغيرات المستقلة. لنعتبر الشكل (٧-٩)، حيث تمثل المنطقة المظللة منطقة المشاهدات لتطبيق انحدار متعدد بمتغيرين مستقلين. والنقطة التي تحيط بها دائرة تقع ضمن مدى كل من المتغيرين المستقلين X_1 و X_2 كلّ على انفراد، ومع ذلك فهي خارجة بوضوح عن المنطقة المشتركة للمشاهدات.

شكل (٧-٩) منطقة مشاهدات لِ X_1 و X_2 معا، بالمقارنة مع مدى X_1 ومدى X_2 كل على انفراد



مراجع ورد ذكرها في النص

[7.1] Box, G.E.P. and Tidwell, P.W. "Transformations of the Independent Variables." *Technometrics* 4(1962), 531 - 50.

[7.2] SAS. *User's Guide: Statistics,* Version 6 edition. Cary, N.C.: SAS Institute, 1987.

مسائل

(٧ـ١) بالإشارة إلى الشكل (٧ـ٤)أ. ماهي، بصورة تقريبية، الزيادة في متوسط الإنتاج، عندما يزداد هطول المطر من 9 إلى 11 بوصة مع بقاء درجة الحرارة ثابتة؟ هل كنت تستطيع الإجابة على هذا السؤال لو كان هطول المطر ودرجة الحرارة يتفاعلان في تأثيريهما على إنتاج المحصول؟

(٧-٢) لنعتبر دالة الاستجابة $E\{Y\} = 25 + 3\,X_1 + 4\,X_2 + 1.5\,X_1X_2$.

أ - ارسم دالة الاستجابة في مقابل X_1 عندما $X_2 = 3$ وعندما $X_2 = 6$. إلى أي حد يظهر تأثير تفاعل X_1 و X_2 على **Y** من هذا البيان؟

ب - خطط مجموعة من منحنيات التساوي لسطح الاستجابة. إلى أي حد يظهر تأثير تفاعل X_1 و X_2 على Y من هذا البيان؟

(٧-٣) لنعتبر دالة الاستجابة $E\{Y\} = 14 + 7\,X_1 - 5\,X_2$.

أ - ارسم دالة الاستجابة في مقابل X_2 عندما $X_1 = 1$ وعندما $X_1 = 4$. كيف يشير البيان إلى أن تأثيرات X_1 و X_2 على Y هي تأثيرات تجميعية؟

ب - خطط مجموعة من منحنيات التساوي لسطح الاستجابة، كيف يشير البيان إلى أن تأثيرات X_1 على X_2 هي تأثيرات تجميعية؟

(٧-٤) اكتب المصفوفة **X** والمتجه β لكل من نموذجي الانحدار التاليين (افترض $i = 1,\ldots, 4$)

أ - $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1} X_{i2} + \varepsilon_i$

ب - $\log Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

(٧-٥) أكتب المصفوفة **X** والمتجه β لكل من نموذجي الانحدار التاليين (افترض $i = 1,\ldots, 4$)

أ - $Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \varepsilon_i$

ب - $\sqrt{Y_i} = \beta_0 + \beta_1 X_{i1} + \beta_2 \log_{10} X_{i2} + \varepsilon_i$

(٧-٦) عرض طالب مايلي: "لا يمكن أن تخفض إضافة متغيرات مستقلة إلى نموذج الانحدار قيمة R^2، وبالتالي ينبغي تضمين كل المتغيرات المستقلة المتوافرة في النموذج". علّق.

(٧-٧) لماذا يخلو إلحاق إشارة بمعامل الانحدار المتعدد R من أي معنى مع أننا نقوم بذلك بالنسبة لمعامل الارتباط البسيط r؟

(٧-٨) **تفضيل صنف.** في دراسة تجريبية، على نطاق ضيق للعلاقة بين درجة تفضيل صنف(Y)وما ينطوي عليه من الرطوبة (X_1) وحلاوة المنتج (X_2)، تم الحصول عل النتائج التالية من تجربة قامت عل أساس التصميم تام العشوائية (البيانات مرمّزة):

$: i$	1	2	3	4	5	6	7	8
$: X_{i1}$	4	4	4	4	6	6	6	6
$: X_{i2}$	2	4	2	4	2	4	2	4
$: Y_i$	64	73	61	76	72	80	71	83

$: i$	9	10	11	12	13	14	15	16
$: X_{i1}$	8	8	8	8	10	10	10	10
$: X_{i2}$	2	4	2	4	2	4	2	4
$: Y_i$	83	89	86	93	88	95	94	100

أ - وفق نموذج الانحدار (7.1) للبيانات. واعرض دالة الانحدار المقدرة. كيف نفسّر b_1 هنا؟

ب - احسب الرواسب وقم بإعداد رسم صندوقي للرواسب. ماهي المعلومات التي يقدمها هذا الرسم ؟

جـ - ارسم الرواسب في مقابل $\hat{Y}$ ، X_1، X_2 وX_1X_2 في رسوم بيانية منفصلة. وقم أيضا بإعداد رسم احتمال طبيعي. حلل الرسومات ولخّص ما توصلت إليه.

د - قم باختبار نقص التوفيق لدالة انحدار من المرتبة الأولى؛ استخدم α=0.01. اعرض البديلين قاعدة القرار، والنتيجة.

(٧-٩) بالإشارة إلى مسألة **تفضيل صنف** (٧-٨) افترض أن نموذج الانحدار (7.1) بحدود خطأ طبيعية مستقلة هو نموذج مناسب.

أ ـ اختبر وجود علاقة انحدار مستخدما مستوى معنوية 0.01. اعرض البدائل، قاعدة القرار، والنتيجة. ماذا يتضمن اختبارك حول β_1 و β_2؟

ب ـ ما هي القيمة -P للاختبار في (أ) ؟

جـ ـ قدّر بصورة مشتركة β_1 و β_2 مستخدما طريقة بونفرّوني ومعامل ثقة عائلي 99 بالمائة. فسّر نتائجك.

(٧-١٠) بالإشارة إلى مسألة **تفضيل صنف** (٧-٨)

أ ـ احسب معامل الانحدار المتعدد R^2 كيف تفسره هنا؟

ب ـ احسب معامل التحديد البسيط r^2 بين Y_i و $\hat{Y}_i$. هل يساوي R^2؟

(٧-١١) بالإشارة إلى مسألة **تفضيل صنف** (٧-٨). افترض أن نموذج الانحدار (7.1) بحدود خطأ طبيعية مستقلة هو نموذج مناسب.

أ ـ أوجد تقديرا بفترة لِـ $E\{Y_h\}$ عندما يكون $X_{h1} = 5$ و $X_{h2} = 4$. استخدم معامل ثقة 99 بالمائة، فسّر تقديرك بفترة هذا.

ب ـ أوجد فترة تنبؤ لمشاهدة جديدة $Y_{h(new)}$ عندما يكون $X_{h1} = 5$ و $X_{h2}=4$. استخدم معامل ثقة 99 بالمائة.

(٧-١٢) **شحن كيماويات**. البيانات التالية، وهي مأخوذة من عشرين شحنة قادمة من كيماويات معبأة في براميل عند وصولها إلى مستودع للبضائع، تبين عدد البراميل في الشحنة (X_1)، الوزن الكلي للشحنة (X_2 ، بمئات الأرطال)، وعدد الدقائق اللازمة لتخزين الشحنة (Y).

10	9	8	7	6	5	4	3	2	1	$: i$
5	16	9	23	5	11	14	5	18	7	$: X_{i1}$
4.76	10.62	7.03	22.07	4.04	10.98	7.03	3.20	16.72	5.11	$: X_{i2}$
44	117	78	203	38	101	93	41	152	58	$: Y_i$

20	19	18	17	16	15	14	13	12	11	$: i$
11	6	21	17	15	8	12	6	12	17	$: X_{i1}$
9.57	3.64	15.21	13.03	13.86	4.60	6.45	3.79	9.51	11.02	$: X_{i2}$
90	39	155	140	127	48	82	50	112	121	$: Y_i$

أ - قم بإعداد رسومات جذع وورقة لأعداد البراميل في الشحنات X_{i1} ولأوزان الشحنات X_{i2} . هل هناك أية مشاهدات قاصية؟ وهل هناك أية ثغرات في البيانات؟

ب - المشاهدات معطاة وفقا لترتيب وصولها. قم بإعداد رسم لكل متغير مستقل في مقابل الزمن. ماذا تبين الرسومات؟

جـ - قم بتوفيق نموذج الانحدار (7.1) لهذه البيانات اكتب دالة الانحدار المقدَّرة. كيف تفسر b_1 و b_2 هنا؟

د - أوجد الرواسب وقم بإعداد رسم صندوقي للرواسب. ماهي المعلومات التي يقدمها الرسم ؟

هـ - ارسم الرواسب في مقابل $\hat{Y}$ ، X_1 ، X_2 وX_1X_2 في رسوم بيانية منفصلة. وقم أيضا بإعداد رسم احتمال طبيعي. حلل الرسومات ولخصّ ما توصلت إليه.

و - قم بإعداد رسم للرواسب مع الزمن. هل هناك أي مؤشر لوجود ارتباط بين حدود الخطأ ؟ ناقش.

(٧-١٣) بالإشارة إلى مسألة **شحن كيماويات** (٧-١٢) افترض أن نموذج الانحدار (7.1) بحدود خطأ طبيعية مستقلة هو نموذج مناسب.

أ - اختبر وجود علاقة انحدار ، مستخدما مستوى معنوية 0.05. اعرض البديلين، قاعدة القرار، والنتيجة. ماذا تتضمن نتيجة اختبارك حول β_1 و β_2 ؟ ما هي القيمة -P لهذا الاختبار؟

ب - قدّر β_1 و β_2 بصورة مشتركة مستخدما طريقة بونفيروني بمعامل ثقة عائلي 95 بالمائة . فسّر نتائجك.

جـ - احسب معامل التحديد المتعدد R^2 . كيف تفسّر هذا القياس هنا؟

(٧-١٤) بالإشارة إلى مسألة **شحن كيماويات** (٧-١٢) افترض أن نموذج الانحدار (7.1) بحدود خطأ طبيعية هو نموذج مناسب.

أ - ترغب الإدارة في الحصول على تقديرات متزامنة بفترة لمتوسط أزمان التخزين لخمس شحنات تقليدية محددة كما يلي:

	1	2	3	4	5
$: X_1$	5	6	10	14	20
$: X_2$	3.20	4.80	7.00	10.00	18.00

أوجد عائلة التقديرات مستخدما معامل ثقة عائلي 95 بالمائة. استخدم حدود من نوع ووركنج - هو تلّنج أو طريقة بونفيرّوني، أيهما أكثر كفاءة.

ب - ومن أجل المشاهدات في المسألة (٧-١٢)، هل تعتبر شحنة من 20 برميلا بوزن يساوي 500 رطلا ضمن مجال النموذج؟ ماذا عن شحنة من 20 برميلا بوزن يساوي 1900 رطلا؟ ادعم نتائجك بإعداد رسم مناسب.

(٧-١٥) بالإشارة إلى مسألة **شحن كيماويات** (٧-١٢). افترض أن نموذج الانحدار (7.1) بحدود خطأ طبيعية هو نموذج مناسب. وفي اليوم أو اليومين التاليين ستصل أربع شحنات منفصلة بياناتها كما يلي:

	1	2	3	4
$: X_1$	9	12	15	18
$: X_2$	7.20	9.00	12.50	16.50

وترغب الإدارة في التنبؤ بأزمنة التخزين لهذه الشحنات بحيث يمكن مقارنة أزمنة التخزين الفعلية بالأزمنة المتنبأ بها وذلك لتحديد ما إذا كانت "خارجة عن المألوف". أوجد التنبؤات التي تحتاجها مستخدما الأسلوب الأكثر كفاءة وبمعامل ثقة عائلي 95 بالمائة.

(٧-١٦) بالإشارة إلى مسألة **شحن كيماويات** (٧-١٢). افترض أن نموذج الانحدار (7.1) بحدود خطأ طبيعية مستقلة هو نموذج مناسب. وسيجري استلام ثلاث شحنات جديدة وفي كل منها $X_{h1} = 7$ و $X_{h2} = 6$.

أ - أوجد 95 بالمائة فترة تنبؤ لمتوسط زمن التخزين لكل من هذه الشحنات.

ب - ردَّ الفترة التي حصلنا عليها في الجزء أ إلى 95 بالمائة فترة تنبؤ لزمن التخزين الكلي للشحنات الثلاث.

(٧-١٧) **ارتياح المريض**. يرغب مدير مستشفى دراسة العلاقة بين ارتياح المريض Y وعمر المريض (X_1 بالسنوات)، شدة المرض (X_2 ، رقم قياسي)، مستوى

القلق (X_3 ، رقم قياسي). وقد اختار عشوائيا 23 مريضا، وجمع البيانات المبينة أدناه، حيث تقترن القيم المرتفعة لـ Y , X_2 و X_3 ، على الترتيب، بارتياح أكبر، زيادة في شدة المرض، وقلق أكبر.

i	1	2	3	4	5	6	7	8	9	10	11	12
X_{i1}	50	36	40	41	28	49	42	45	52	29	29	43
X_{i2}	51	46	48	44	43	54	50	48	62	50	48	53
X_{i3}	2.3	2.3	2.2	1.8	1.8	2.9	2.2	2.4	2.9	2.1	2.4	2.4
Y_i	48	57	66	70	89	36	46	54	26	77	89	67

i	13	14	15	16	17	18	19	20	21	22	23
X_{i1}	38	34	53	36	33	29	33	55	29	44	43
X_{i2}	55	51	54	49	56	46	49	51	52	58	50
X_{i3}	2.2	2.3	2.2	2.0	2.5	1.9	2.1	2.4	2.3	2.9	2.3
Y_i	47	51	57	66	79	88	60	49	77	52	60

أ ـ قم بإعداد رسم جذع وورقة لكل من المتغيرات المستقلة. هل تكشف هذه الرسومات عن أية جوانب تستحق الملاحظة؟

ب ـ وفق النموذج (7.5) بثلاثة متغيرات مستقلة لهذه البيانات واكتب دالة الانحدار المقدَّرة. كيف تفسّر b_2 هنا؟.

جـ ـ أوجد الرواسب وقم بإعداد رسم صندوقي للرواسب. هل تبدو هناك أية مشاهدات قاصية؟

د ـ ارسم الرواسب في مقابل $\hat{Y}$ ، وفي مقابل كل من المتغيرات المستقلة، وفي مقابل كل تفاعل بين عاملين في رسوم بيانية منفصلة. وقم أيضا بإعداد رسم احتمال طبيعي. حلل رسوماتك ولخصّ ماتوصلت إليه.

هـ ـ هل تستطيع القيام باختبار نقص التوفيق هنا؟

(٧-١٨) بالإشارة إلى مسألة **ارتياح المريض** (٧-١٧). افترض أن نموذج الانحدار (7.5) لثلاثة متغيرات مستقلة، مع حدود خطأ طبيعية مستقلة، هو نموذج مناسب.

أ - اختبر ما إذا كانت توجد علاقة انحدار؛ استخدم مستوى معنوية 0.01. اعرض البديلين، قاعدة القرار، والنتيجة. ماذا يتضمن اختبارك حول β_1، β_2 و β_3؟ ما هي القيمة -P للاختبار؟

ب - أوجد تقديرات مشتركة بفترة لِـ β_1، β_2 و β_3 مستخدما معامل ثقة عائلي 90 بالمائة. فسّر نتائجك.

جـ - احسب معامل الارتباط المتعدد. إلام يشير هنا؟.

(٧-١٩) بالاشارة إلى مسألة **ارتياح المريض** (٧-١٧). افترض أن نموذج الانحدار (7.5) لثلاثة متغيرات مستقلة مع حدود خطأ طبيعية مستقلة، هو نموذج مناسب.

أ - أوجد تقديرا بفترة لمتوسط الارتياح عندما يكون $X_{h1} = 35$ ، $X_{h2} = 45$ و$X_{h3} = 2.2$. استخدم معامل ثقة 90 بالمئة. فسّر فترة الثقة التي حصلت عليها.

ب - أوجد فترة تنبؤ لارتياح مريض جديد عندما يكون $X_{h1} = 35$ ، $X_{h2} = 45$ و$X_{h3} = 2.2$. استخدم معامل ثقة 90 بالمائة. فسّر فترة التنبؤ التي حصلت عليها.

(٧-٢٠) **رواتب المختصين في الرياضيات**. رغب باحث في مؤسسة علمية في تثمين العلاقة بين الرواتب السنوية لباحثين في الرياضيات من مستوى متوسط ومتقدم (Y ، بآلاف الدولارات) ورقم قياسي يعبر عن نوعية المنشورات (X_1)، عدد سنوات الخبرة (X_2)، ورقم قياسي يعبر عن النجاح في الحصول على دعم منحة (X_3). وفيما يلي البيانات لعينة من 24 باحثا في الرياضيات من مستويات متوسطة ومتقدمة.

i:	1	2	3	4	5	6	7	8	9	10	11	12
X_{i1}:	3.5	5.3	5.1	5.8	4.2	6.0	6.8	5.5	3.1	7.2	4.5	4.9
X_{i2}:	9	20	18	33	31	13	25	30	5	47	25	11
X_{i3}:	6.1	6.4	7.4	6.7	7.5	5.9	6.0	4.0	5.8	8.3	5.0	6.4
Y_i:	33.2	40.3	38.7	46.8	41.4	37.5	39.0	40.7	30.1	52.9	38.2	31.8

i:	13	14	15	16	17	18	19	20	21	22	23	24
X_{i1}:	8.0	6.5	6.6	3.7	6.2	7.0	4.0	4.5	5.9	5.6	4.8	3.9
X_{i2}:	23	35	39	21	7	40	35	23	33	27	34	15
X_{i3}:	7.6	7.0	5.0	4.4	5.5	7.0	6.0	3.5	4.9	4.3	8.0	5.0
Y_i:	43.3	44.1	42.8	33.6	34.2	48.0	38.0	35.9	40.4	36.8	45.2	35.1

أ ـ قم بإعداد رسم جذع وورقة لكل من المتغيرات المستقلة. ماهي المعلومات التي تقدمها هذه الرسومات؟

ب ـ قم بتوفيق نموذج الانحدار (7.5) بثلاث متغيرات مستقلة لهذه البيانات. اكتب دالة الانحدار المقدرة.

جـ ـ أوجد الرواسب وقم بإعداد رسم صندوقي للرواسب. هل يبدو التوزيع متناظرا إلى حد مقبول.

د ـ ارسم الرواسب في مقابل $\hat{Y}$، وفي مقابل كل من المتغيرات المستقلة وكل من التفاعلات بين عاملين وذلك في رسوم بيانية منفصلة. وقم أيضا بإعداد رسم احتمال طبيعي. حلل رسوماتك ولخص ماتوصلت إليه.

هـ ـ هل يمكنك القيام باختبار نقص التوفيق هنا؟

(٧ـ٢١) بالإشارة إلى مسألة رواتب **المختصين في الرياضيات** (٧ـ٢٠). افترض أن نموذج الانحدار (7.5) لثلاثة متغيرات مستقلة مع حدود خطأ طبيعية مستقلة، هو نموذج مناسب.

أ ـ اختبر ما إذا كانت توجد علاقة انحدار؛ استخدم $\alpha = 0.05$، اعرض البديلين، قاعدة القرار، والنتيجة. ماذا يتضمن اختبارك حول β_1، β_2 وβ_3 ما هي القيمة -P للاختبار؟

ب ـ قدّر بصورة مشتركة β_1، β_2 و β_3 باستخدم طريقة بونفرّوني مستخدما معامل ثقة عائلي 95 بالمائة. فسّر نتائجك.

جـ ـ احسب R^2 وفسّر هذا المقياس.

(٧ـ٢٢) بالإشارة إلى مسألة **رواتب المختصين في الرياضيات** (٧ـ٢٠). افترض أن نموذج الانحدار (7.5) لثلاثة متغيرات مستقلة مع حدود خطأ طبيعية مستقلة، هو نموذج مناسب. ويرغب الباحث في الحصول على تقديرات متزامنة بفترة لمتوسطات مستويات الرواتب لأربعة من الباحثين الرياضيين التقليديين مواصفاتهم كما يلي:

	1	2	3	4
$: X_1$	5.0	6.0	4.0	7.0
$: X_2$	20	30	10	50
$: X_3$	5.0	6.0	4.0	7.0

أوجد العائلة من التقديرات مستخدما معامل ثقة عائلي 95 بالمائة. استخدم الطريقة الأكثر كفاءة.

(٧-٢٣) بالإشارة إلى مسألة **رواتب المختصين في الرياضيات** (٧-٢٠). افترض أن نموذج الانحدار (7.5) لثلاثة متغيرات مستقلة مع حدود خطأ طبيعية مستقلة، هو نموذج مناسب. ولم يقدم ثلاثة من الباحثين الرياضيين الذين شملتهم الدراسة أية معلومات عن الرواتب، وكانت مواصفاتهم كما يلي:

	1	2	3
$: X_1$	5.4	6.2	6.4
$: X_2$	17	12	21
$: X_3$	6.0	5.8	6.1

طوّر فترات تنبؤ منفصلة للرواتب السنوية لهؤلاء الرياضيين مستخدما في كل حالة معامل ثقة 95 بالمائة. هل يمكن التنبؤ برواتب الرياضيين الثلاثة هؤلاء بدقة مقبولة ؟

تمارين

(٧-٢٤) لكل من نماذج الانحدار التالية، أشر إلى ما إذا كان النموذج هو نموذج انحدار خطّي عام. واذا لم يكن كذلك، اذكر ما إذا كان يمكن التعبير عنه في الصيغة (7.7) باستخدام تحويل مناسب:

أ) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \log_{10} X_{i2} + \beta_3 X_{i1}^2 + \varepsilon_i$

ب) $Y_i = \varepsilon_i \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2)$

جـ) $Y_i = \beta_0 + \log_{10}(\beta_1 X_{i1}) + \beta_2 X_{i2} + \varepsilon_i$

د) $Y_i = \beta_0 \exp(\beta_1 X_{i1}) + \varepsilon_i$

هـ) $Y_i = [1 + \exp(\beta_0 + \beta_1 X_{i1} + \varepsilon_i)]^{-1}$

(٧-٢٥) (في حاجة إلى حساب التفاضل) لنعتبر نموذج الانحدار المتعدد:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad i = 1,\dots, n$$

حيث المقادير ε_i غير مرتبطة، $E\{\varepsilon_i\} = 0$ و $\sigma^2\{\varepsilon_i\} = \sigma^2$.

أ ـ استنبط مقدِّرات المربعات الدنيا لِـ β_1 و β_2.

ب ـ مفترضا أن المقادير ε_i هي متغيرات عشوائية طبيعية مستقلة. اكتب دالة الإمكانية، وأوجد مقدِّرات الإمكانية العظمى لِـ β_1 و β_2. هل تتطابق هذه المقدِّرات مع مقدِّرات المربعات الدنيا ؟

(٧-٢٦) (في حاجة إلى حساب التفاضل) لنعتبر نموذج الانحدار المتعدد:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \varepsilon_i \qquad i = 1, ..., n$$

حيث المقادير ε_i مستقلة $N(0, \sigma^2)$. استنبط المعادلات الناظمية بطريقة المربعات الدنيا. هل ستقدم هذه المعادلات مقدِّرات لمعاملات الانحدار مطابقة لمقدِّرات الإمكانية العظمى؟

(٧-٢٧) أراد محلل توفيق نموذج الانحدار

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad i = 1, ..., n$$

بطريقة المربعات الدنيا حيث كان من المعروف أن $\beta_2 = 4$ كيف يمكن للمحلل الحصول على التوفيق المرغوب مستخدما برنامج حاسب خاص بالانحدار المتعدد؟

(٧-٢٨) من أجل نموذج الانحدار (7.1)، بيِّن أن معامل التحديد البسيط r^2 بين Y_i و $\hat{Y}_i$ يساوي معامل التحديد المتعدد R^2.

(٧-٢٩) في دراسة انحدار على نطاق ضيق، حصلنا على البيانات التالية:

6	5	4	3	2	1	$: i$
8	21	3	16	4	7	$: X_1$
31	5	49	7	41	33	$: X_2$
55	91	28	75	33	42	$: X_3$

افترض أن نموذج الانحدار (7.1) مع حدود خطأ طبيعية مستقلة هو نموذج مناسب مستخدما طرق المصفوفات، أوجد

(أ) **b** ؛ (ب) **e** ؛ (جـ) **H** ؛ (د) SSR؛ (هـ) $s^2\{\mathbf{b}\}$ (و) $\hat{Y}_h$ عندما يكون $X_{h1} = 10$ و$X_{h2} = 30$ ، (ز) $s^2\{\hat{Y}_h\}$ عندما يكون $X_{h1} = 10$ و$X_{h2} = 30$.

مشاريع

(٧-٣٠) بالإشارة إلى مجموعة البيانات **SMSA**. طُلب منك تقويم نموذجين بديلين للتنبؤ بعدد الأطباء الممارسين Y في **SMSA**.ويتضمن النموذج الأول المقترح كمتغيرات مستقلة: العدد الكلي للسكان (X_1)، مساحة الأرض (X_2)، والدخل الشخصي الإجمالي (X_3). والنموذج الثاني المقترح يتضمن كمتغيرات مستقلة: كثافة السكان (X_1 ، العدد الكلي للسكان مقسوما على مساحة الأرض)، النسبة المئوية للسكان في المدن المركزية (X_2)، والدخل الشخصي الإجمالي (X_3).

أ - جهّز بإعداد رسم جذع وورقة لكل من المتغيرات المستقلة. ماهي المعلومات التي تستحق الذكر التي تقدمها الرسومات ؟

ب - لكل من النموذجين المقترحين، وفّق نموذج الانحدار من المرتبة الأولى (7.5) بثلاثة متغيرات مستقلة.

جـ - احسب R^2 لكل نموذج. هل أحد النموذجين مفضل تفضيلا واضحا على الآخر وفقا لهذا المقياس ؟

د - لكل من النموذجين، أوجد الرواسب وارسمها في مقابل $\hat{Y}$، وفي مقابل كل من المتغيرات المستقلة، وكل من التفاعلات ذات العاملين. قم أيضا بإعداد رسم احتمال طبيعي لكل من النموذجين اللذين قمت بتوفيقهما. حلل رسوماتك ولخصّ ماتوصلت إليه. هل أحد النموذجين مفضل بوضوح من حيث المصداقية؟

(٧-٣١) بالإشارة إلى مجموعة البيانات **SMSA**

أ - لكل منطقة جغرافية، إحدر عدد الجرائم الخطرة في **SMSA** (Y) على الكثافة السكانية (X_1 العدد الكلي للسكان مقسوما على مساحة الأرض)، الدخل الشخصي الإجمالي (X_2)، النسبة المئوية لخريجي الثانوية (X_3). استخدم نموذج الانحدار من المرتبة الأولى (7.5) بثلاثة متغيرات مستقلة. اكتب دوال الإنحدار المقدّرة.

ب ـ هل دوال الانحدار المقدَّرة متشابهة بالنسبة للمناطق الأربع؟ ناقش.

جـ ـ احسب **MSE** وR^2 لكل منطقة. هل هذان المقياسان متماثلان بالنسبة للمناطق الأربع؟ ناقش.

د ـ أوجد الرواسب لكل نموذج قمت بتوفيقه، واعرض رسما صندوقيا للرواسب في كل من النموذجين. حلل رسوماتك واعرض ماتوصلت إليه.

(٧-٣٢) بالإشارة إلى مجموعة البيانات **SENIC**. اقتُرح نموذجان للتنبؤ بمتوسط مدة إقامة المريض في المستشفى Y. ويستخدم النموذج الأول كمتغيرات مستقلة: العمر (X_1)، خطورة الإصابة (X_2)، والخدمات والتسهيلات المتوافرة (X_3). ويستخدم النموذج الثاني كمتغيرات مستقلة: عدد الأسِرَّة(X_1)، خطورة الإصابة (X_2)، والخدمات والتسهيلات المتوافرة (X_3).

أ ـ جهِّز بإعداد رسم جذع وورقة لكل من المتغيرات المستقلة. ماهي المعلومات التي تقدمها هذه الرسومات ؟.

ب ـ لكل من النموذجين المقترحين، وفِّق نموذج الانحدار (7.5) من المرتبة الأولى بثلاثة متغيرات مستقلة.

جـ ـ احسب R^2 لكل نموذج. هل أحد النموذجين مفضل بوضوح وفقا لهذا المقياس؟

د ـ لكل نموذج، أوجد الرواسب وارسمها في مقابل $\hat{Y}$، وفي مقابل كل من المتغيرات المستقلة، وكل من التفاعلات ذات العاملين. وقم أيضا بإعداد رسم احتمال طبيعي للرواسب في كل من النموذجين اللذين قمت بتوفيقهما. حلل رسوماتك واعرض ماتوصلت إليه. هل أحد النموذجين مفضل بوضوح من حيث المصداقية؟

(٧-٣٣) بالإشارة إلى مجموعة البيانات **SENIC**.

أ ـ إحدر، لكل منطقة جغرافية، خطورة الإصابة Y على المتغيرات المستقلة، العمر (X_1)، نسبة الزرع الروتيني (X_2)، معدل الإحصاءات اليومية لعدد المرضى (X_3)، والتسهيلات والخدمات المتوافرة (X_4).

استخدم نموذج الانحدار من المرتبة الأولى (7.5) بأربعة متغيرات مستقلة. اكتب دوال الانحدار المقدَّرة.

ب - هل دوال الانحدار المقدَّرة متماثلة في المناطق الجغرافية الأربع؟ ناقش.

جـ - احسب **MSE** وR^2 لكل منطقة. هل هذان القياسان متشابهان في المناطق الأربع ؟ ناقش.

د - أوجد الرواسب لكل نموذج قمت بتوفيقه. واعرض رسما صندوقيا للرواسب في كل من النموذجين حلل رسوماتك واعرض ماتوصلت إليه.

الفصل الثامن

الانحدار المتعدد - II

نتابع في هذا الفصل، عدّة مواضيع خاصة بالانحدار المتعدد. فندرس أولا مجاميع المربعات الإضافية، وهي مفيدة للقيام باختبارات متنوعة حول معاملات الانحدار. ثم نتابع نسخة من نموذج الانحدار المتعدد تكون القياسات فيها معيارية ونقدم الخطّية المتعددة وهو شرط تكون معه المتغيرات المستقلة مرتبطة ارتباطا عاليا. ونقدم، أخيرا، الاختبار الخطّي العام في صيغة مصفوفية.

(٨ -١) مجاميع المربعات الإضافية

أفكار أساسية

يقيس مجموع مربعات إضافي التخفيض الهامشي في مجموع مربعات الخطأ عند إضافة متغير أو عدة متغيرات مستقلة إلى نموذج الانحدار، علما أن المتغيرات الأخرى كانت من حينها ضمن النموذج. وبصورة مكافئة ، يمكن النظر إلى مجموع مربعات إضافي كقياس للزيادة الهامشية في مجموع مربعات الانحدار عند إضافة متغير أو عدّة متغيرات مستقلة إلى نموذج الانحدار.

وسنستخدم أولا مثالا لتوضيح هذه الأفكار، ثم نقدم تعاريف لمجاميع المربعات الإضافية. ونناقش استخدامات متنوعة لمجاميع المربعات الإضافية في اختبارات حول معاملات الانحدار.

مثال. يتضمن الجدول (٨-١)أ بيانات من دراسة للعلاقة بين مقدار الشحوم في الجسم Y وعدة متغيرات تفسيرية مستقلة. وتقوم الدراسة على عينة من إناث يتمتعن بصحة جيدة وأعمارهن من 25 إلى 34 سنة، والمتغيرات المستقلة الممكنة هي سماكة الجلد في عضلة مؤخر العضد ثلاثية الرؤوس (X_1)، محيط الفخذ (X_2)، ومحيط منتصف الذراع (X_3).

جدول (٨-١) البيانات ومصفوفة الارتباط بين المتغيرات X لمثال شحوم الجسم.

(أ) بيانات شخصية

الشخص i	سماكة الجلد في عضلة مؤخر العضد X_{i1}	محيط الفخذ X_{i2}	محيط منتصف الذراع X_{i3}	شحوم الجسم Y_i
1	19.5	43.1	29.1	11.9
2	24.7	49.8	28.2	22.8
3	30.7	51.9	37.0	18.7
4	29.8	54.3	31.1	20.1
5	19.1	42.2	30.9	12.9
6	25.6	53.9	23.7	21.7
7	31.4	58.5	27.6	27.1
8	27.9	52.1	30.6	25.4
9	22.1	49.9	23.2	21.3
10	25.5	53.5	24.8	19.3
11	31.1	56.6	30.0	25.4
12	30.4	56.7	28.3	27.2
13	18.7	46.5	23.0	11.7
14	19.7	44.2	28.6	17.8
15	14.6	42.7	21.3	12.8
16	29.5	54.4	30.1	23.9
17	27.7	55.3	25.7	22.6
18	30.2	58.6	24.6	25.4
19	22.7	48.2	27.1	14.8
20	25.2	51.0	27.5	21.1

(ب) مصفوفة الارتباط للمتغيرات X

$$\mathbf{r}_{XX} = \begin{bmatrix} 1.0 & 0.92 & 0.46 \\ 0.92 & 1.0 & 0.08 \\ 0.46 & 0.08 & 1.0 \end{bmatrix}$$

ويتضمن الجدول (٨-٢) نتائج الانحدار عندما نحدر شحوم الجسم Y على (١) سماكة الجلد في عضلة مؤخر العضد (X_1) بمفرده ، (٢) على محيط الفخذ (X_2) بمفرده، (٣) على X_1 و X_2 فقط ، و(٤) على المتغيرات المستقلة الثلاثة جميعها. ولاقتفاء نموذج الانحدار الذي جرى توفيقه، سنعدّل رموزنا قليلا. فمجموع مربعات الانحدار عندما يكون X_1 فقط في النموذج هو، وفقا للجدول (٨-٢)أ، 352.27 وسنرمز لمجموع المربعات هذا بـ $SSR(X_1)$ وسنرمز لمجموع مربعات الخطأ في هذه النموذج بـ $SSE(X_1)$ ووفقا للجدول (٨-٢)أ لدينا $SSE(X_1) = 143.12$.

جدول (٨-٢) نتائج الانحدار لعدّة نماذج جرى توفيقها في مثال شحوم الجسم.

(أ) انحدار Y على X_1

$$\hat{Y} = -1.496 + 0.8572 X_1$$

مصدر التغير	SS	df	MS
الانحدار	352.27	1	352.27
الخطأ	143.12	18	7.95
المجموع	495.39	19	

المتغير	معامل الانحدار المقدّر	الانحراف المعياري المقدر	t^*
X_1	$b_1 = 0.8572$	$s\{b_1\} = 0.1288$	66.6

(ب) انحدار Y على X_2

$$\hat{Y} = -23.634 + 0.8565 X_2$$

مصدر التغير	SS	df	MS
الانحدار	381.97	1	381.97
الخطأ	113.42	18	6.30
المجموع	495.39	19	

المتغير	معامل الانحدار المقدّر	الانحراف المعياري المقدّر	t^*
X_2	$b_2 = 0.8565$	$s\{b_2\} = 0.1100$	7.79

(جـ) انحدار Y على X_1 و X_2

$$\hat{Y} = -19.174 + 0.2224 X_1 + 0.6594 X_2$$

مصدر التغير	SS	df	MS
الانحدار	385.44	2	192.72
الخطأ	109.95	17	6.47
المجموع	495.39	19	

المتغير	معامل الانحدار المقدّر	الانحراف المعياري المقدّر	t^*
X_1	$b_1 = 0.2224$	$s\{b_1\} = 0.3034$	0.73
X_2	$b_2 = 0.6594$	$s\{b_2\} = 0.2912$	2.26

تتمة الجدول ٨-٢

(د) انحدار Y على X_1، X_2 وX_3

$$\hat{Y} = 117.08 + 4.334X_1 - 2.857X_2 - 2.186X_3$$

مصدر التغير	SS	df	MS
الانحدار	396.98	3	132.33
الخطأ	98.41	16	6.15
المجموع	495.39	19	

المتغير	معامل الانحدار المقدّر	الانحراف المعياري المقدّر	t^*
X_1	$b_1 = 4.334$	$s\{b_1\} = 3.016$	1.44
X_2	$b_2 = -2.857$	$s\{b_2\} = 2.582$	-1.11
X_3	$b_3 = -2.186$	$s\{b_3\} = 1.596$	-1.37

وبصورة مشابهة، يشير الجدول (٨-٢)جـ، عند وجود X_1 وX_2 في نموذج الانحدار، إلى أن مجموع مربعات الانحدار $SSR(X_1, X_2) = 385.44$ وأن مجموع مربعات الخطأ $SSE(X_1, X_2) = 109.95$.

ونلاحظ أن مجموع مربعات الخطأ عند وجود X_1 و X_2 في النموذج، وهو $SSE(X_1, X_2) = 109.95$ أصغر من قيمته عندما يتضمن النموذج X_1 فقط، إذ يكون عندئذ $SSE(X_1) = 143.12$. ويدعى الفرق بمجموع مربعات إضافي وسنرمز له بـ $SSR(X_2 | X_1)$:

$$SSR(X_2 | X_1) = SSE(X_1) - SSE(X_1, X_2) = 143.12 - 109.95 = 33.17$$

وهذا التخفيض في مجموع مربعات الخطأ كان نتيجة لإضافة X_2 إلى نموذج الانحدار حيث كانت X_1 من حينها مشمولة في النموذج. وهكذا يقيس مجموع المربعات الإضافي $SSR(X_2 | X_1)$ التأثير الهامشي لإضافة X_2 إلى نموذج انحدار كان يقتصر على X_1. ويعكس الرمز $SSR(X_2 | X_1)$ هذا التخفيض الإضافي في مجموع مربعات الخطأ الذي يترافق مع X_2 علما أن X_1 كانت من حينها مشمولة في النموذج.

ويقيس مجموع المربعات الإضافي $SSR(X_2 | X_1)$ بصورة مكافئة، الزيادة الهامشية في مجموع مربعات الانحدار:

$$SSR(X_2 | X_1) = SSR(X_1, X_2) - SSR(X_1) = 385.44 - 352.27 = 33.17$$

وسبب تكافؤ التخفيض الهامشي في مجموع مربعات الخطأ والزيادة الهامشية في مجموع مربعات الانحدار هو المطابقة الأساسية في تحليل التباين (3.48a):

$$SSTO = SSR + SSE$$

وبما أن $SSTO$ يقيس متغيرية المشاهدات Y_i، وبالتالي لايعتمد على نموذج الانحدار الذي جرى توفيقه، فإن أي تخفيض في SSE يتضمن زيادة مطابقة في SSR.

ويمكن دراسة مجاميع المربعات الأخرى، مثل التأثير الهامشي لإضافة X_3 إلى نموذج انحدار يتضمن من حينه X_1 و X_2. ونجد من الجدولين (٨-٢)جـ و(٨-٢)د أن:

$$SSR(X_3 \mid X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3)$$
$$= 109.95 - 98.41 = 11.54$$

أو، بصورة مكافئة:

$$SSR(X_3 \mid X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2)$$
$$= 396.98 - 385.44 = 11.54$$

ويمكن دراسة حتى التأثير الهامشي لإضافة عدة متغيرات، مثل إضافة كل من X_2 و X_3 إلى نموذج انحدار يتضمن X_1 (انظر الجدولين (٨-٢)أ و(٨-٢)د):

$$SSR(X_2, X_3 \mid X_1) = SSE(X_1) - SSE(X_1, X_2, X_3)$$
$$= 143.12 - 98.41 = 44.71$$

أو بصورة مكافئة:

$$SSR(X_2, X_3 \mid X_1) = SSR(X_1, X_2, X_3) - SSR(X_1)$$
$$= 396.98 - 352.27 = 44.71$$

تعاريف

نجمّع الآن تعاريفنا السابقة لمجاميع المربعات الإضافية ونقدم بعض التعاريف الإضافية. وكما لاحظنا سابقا، فإن مجموع المربعات الإضافي ينطوي دائما على الفرق بين مجموع مربعات الخطأ في نموذج الانحدار المتضمن المتغير، أو متغيرات X ، ومجموع مربعات الخطأ للنموذج نفسه بعد إضافة متغير أو متغيرات X إليه. وبصورة مكافئة، ينطوي مجموع المربعات الإضافي على الفرق بين مجموعي مربعات الانحدار المقابلين.

وهكذا نجد:

$$SSR(X_1 \mid X_2) = SSE(X_2) - SSE(X_1, X_2) \qquad (8.1a)$$

أو، بصورة مكافئة:

$$SSR(X_1 \mid X_2) = SSR(X_1, X_2) - SSR(X_2) \qquad (8.1b)$$

وإذا كان X_2 هو المتغير الإضافي ، فلدينا:

$$SSR(X_2 \mid X_1) = SSE(X_1) - SSE(X_1, X_2) \qquad (8.2a)$$

أو، بصورة مكافئة:

$$SSR(X_2 \mid X_1) = SSR(X_1, X_2) - SSR(X_1) \quad (8.2b)$$

ويمكن بصورة مباشرة التعميم إلى حالـة ثلاثـة متغيرات أو أكثر. وعلى سبيل المثال، لدينا:

$$SSR(X_3 \mid X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3) \quad (8.3a)$$

$$SSR(X_3 \mid X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2) \quad (8.3b)$$

و:

$$SSR(X_2, X_3 \mid X_1) = SSE(X_1) - SSE(X_1, X_2, X_3) \quad (8.4a)$$

$$SSR(X_2, X_3 \mid X_1) = SSR(X_1, X_2, X_3) - SSR(X_1) \quad (8.4b)$$

تفكيك SSR إلى مجاميع مربعات إضافية

في الانحدار المتعدد، وخلافا للانحـدار البسيط، يمكن الحصول على تفكيكات متنوعة لمجموع مربعات الانحدار SSR إلى مجاميع مربعات إضافية لنعتبر حالة متغيرين مستقلين. ونبدأ بالمطابقة (3.48a) لمتغير X_1:

$$SSTO = SSR(X_1) + SSE(X_1) \quad (8.5)$$

حيث تبين الرموز الآن، بوضوح، أن X_1 هو المتغير X في النموذج، وإذا وضعنا بدلا من $SSE(X_1)$ مكافئه من (8.2a)، نجد:

$$SSTO = SSR(X_1) + SSR(X_2 \mid X_1) + SSE(X_1, X_2) \quad (8.6)$$

وفي حالة الانحدار المتعدد بمتغيرين مستقلين، لدينا المطابقة نفسها كما في (8.5) الخاصة بمتغير مستقل واحد، ونعني:

$$SSTO = SSR(X_1, X_2) + SSE(X_1, X_2) \quad (8.7)$$

وبحل (8.7) كمعادلة في المجهول $SSE(X_1, X_2)$ واستخدام العبارة الناتجة في (8.6) نجد:

$$SSR(X_1, X_2) = SSR(X_1) + SSR(X_2 \mid X_1) \quad (8.8)$$

وهكذا نكون قد فككنا مجموع مربعات الانحدار $SSR(X_1, X_2)$ إلى مركبتين هامشيتين: (١) $SSR(X_1)$ تقيس المساهمة الناتجة عن وجود X_1 بمفرده في النموذج و(٢) $SSR(X_2 \mid X_1)$ وهي تقيس المساهمة الإضافية الناتجة عن ضم X_2 إلى النموذج علما أن X_1 موجود من حينه في النموذج.

وبالطبع فإن ترتيب المتغيرات X كيفي، إذ يمكننا هنا الحصول أيضا على التفكيك:

$$SSR(X_1, X_2) = SSR(X_2) + SSR(X_1 \mid X_2) \quad (8.9)$$

ونبين في الشكل (٨-١) تمثيلين تخطيطيين لتفكيكي $SSR(X_1, X_2)$ في مثال شحوم الجسم. ويمثل المستطيل الكلي على اليسار $SSTO$ ويقدم التفكيك (8.9).

والمركبة غير المظللة في هذا المستطيل هي $SSR(X_2)$ وتمثل المنطقة المظللة بقسميها $SSE\ (X_2)$. وهذه المنطقة الأخيرة هي بدورها مركبة من مجموع المربعات الإضافي $SSR(X_1 \mid X_2)$ ومجموع مربعات الخطأ $SSE(X_1 , X_2)$ وهو مجموع مربعات الخطأ الناتج عند وجود X_1 وX_2 كليهما في النموذج. وبصورة مشابهة، يبين المستطيل على اليمين في الشكل (٨-١) التفكيك (8.8). ونلاحظ، في الحالتين، كيف يمكن النظر إلى مجموع المربعات الإضافي كتخفيض في مجموع مربعات الخطأ، أو كزيادة في مجموع مربعات الانحدار عند إضافة المتغير الثاني إلى نموذج الانحدار.

وعندما يتضمن نموذج الانحدار ثلاثة متغيرات X، فيمكن الحصول على تفكيكات متنوعة لـ $SSR(X_1 , X_2, X_3)$. ونوضح فيما يلي ثلاثا منها:

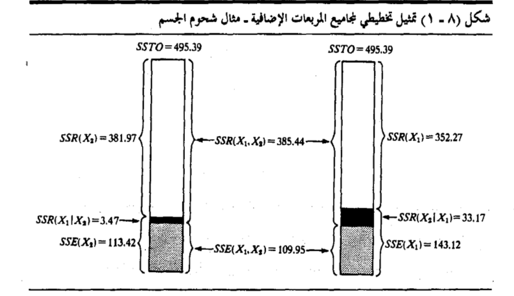
$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2 \mid X_1) + SSR(X_3 \mid X_1 , X_2) \quad (8.10a)$$

$$SSR(X_1, X_2, X_3) = SSR(X_2) + SSR(X_3 \mid X_2) + SSR(X_1 \mid X_2 , X_3) \quad (8.10b)$$

$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2, X_3 \mid X_1) \quad (8.10c)$$

ومن الواضح أن عدد التفكيكات الممكنة يصبح واسعا عندما يزداد عدد المتغيرات X في النموذج.

شكل (٨ ـ ١) تمثيل تخطيطي لمجاميع المربعات الإضافية ـ مثال شحوم الجسم



جدول تباين يتضمن تفكيك SSR

يمكن إقامة جداول التباين المتضمنة لتفكيكات مجموع مربعات الانحدار إلى مجاميع

مربعات إضافية. ويتضمن الجدول (٨-٣) جدول التحاين لتفكيك ممكن في حالة ثلاثة متغيرات مستقلة، ويتضمن الجدول (٨-٤) التفكيك نفسه لمثال شحوم الجسم.

جدول (٨ -٣) مثال جدول تحاين مع تفكيك SSR في حالة ثلاثة متغيرات مستقلة.

مصدر التغير	SS	df	MS
الانحدار	$SSR(X_1, X_2, X_3)$	3	$MSR(X_1, X_2, X_3)$
X_1	$SSR(X_1)$	1	$MSR(X_1)$
$X_2 \mid X_1$	$SSR(X_2 \mid X_1)$	1	$MSR(X_2 \mid X_1)$
$X_3 \mid X_1, X_2$	$SSR(X_3 \mid X_1, X_2)$	1	$MSR(X_3 \mid X_1, X_2)$
الخطأ	$SSE(X_1, X_2, X_3)$	$n - 4$	$MSE(X_1, X_2, X_3)$
المجموع	$SSTO$	$n - 1$	

جدول (٨ -٤) جدول تحاين مع تفكيك SSR - مثال شحوم الجسم بثلاثة متغيرات مستقلة.

مصدر التغير	SS	df	MS
الانحدار	$SSR(X_1, X_2, X_3) = 396.98$	3	132.33
X_1	$SSR(X_1) = 352.27$	1	352.27
$X_2 \mid X_1$	$SSR(X_2 \mid X_1) = 33.17$	1	33.17
$X_3 \mid X_1, X_2$	$SSR(X_3 \mid X_1, X_2) = 11.54$	1	11.54
الخطأ	$SSE(X_1, X_2, X_3) = 98.41$	16	6.15
المجموع	$SSTO = 495.39$	19	.

لاحظ أن كل مجموع مربعات إضافي ينطوي على متغير إضافي واحد قد اقترن بدرجة واحدة من الحرية، واقترنت مجاميع المربعات الإضافية المنطوية على متغيرين إضافيين مثل $SSR(X_2, X_3 \mid X_1)$ بدرجتين من الحرية. وينتج ذلك بسبب إمكانية التعبير عن مجموع مربعات كهذا كمجموع مجموعي مربعات إضافيين، يقترن كل منهما بدرجة واحدة من الحرية وعلى سبيل المثال، لدينا من تعريف مجاميع المربعات الإضافية:

$$SSR(X_2, X_3 \mid X_1) = SSR(X_2 \mid X_1) + SSR(X_3 \mid X_1, X_2) \qquad (8.11)$$

ويقدم عدد من حزم الحاسب الخاصة بالانحدار تفكيكات لـِ SSR إلى مجاميع مربعات إضافية كل منها بدرجة واحدة من الحرية، ويكون ذلك، عادة، بالترتيب نفسه الذي أُدخلت فيه المتغيرات X إلى النموذج. وهكذا، إذا أدخلت المتغيرات المستقلة بالترتيب X_1، X_2 وX_3 فإن مجاميع المربعات الإضافية المعطاة في المخرجات تكون:

$$SSR(X_1)$$
$$SSR(X_2 \mid X_1)$$
$$SSR(X_3 \mid X_1, X_2)$$

وإذا رغبنا بمجموع مربعات إضافي ينطوي على عدة متغيرات مستقلة. فيمكن الحصول عليه بجمع مايناسب من مجاميع المربعات الإضافية بدرجة واحدة من الحرية. وعلى سبيل المثال، للحصول على $SSR(X_2, X_3 \mid X_1)$ في التوضيح السابق، يمكن استخدام (8.11) فنجمع، ببساطة، $SSR(X_2 \mid X_1)$ و $SSR(X_3 \mid X_1, X_2)$.

وإذا رغبنا بمجموع المربعات الإضافي $SSR(X_1, X_3 \mid X_2)$ من حزمة حاسب تقدم مجاميع مربعات إضافية بدرجة واحدة من الحرية بالترتيب الذي أدخلت فيه المتغيرات X، فسنحتاج إلى أن تكون المتغيرات المستقلة قد أدخلت بالترتيب X_2 ثم X_1 ثم X_3 أو X_2 ثم X_3 ثم X_1. والترتيب الأول يعطي:

$$SSR(X_2)$$
$$SSR(X_1 \mid X_2)$$
$$SSR(X_3 \mid X_1, X_2)$$

ومجموع مجموعي المربعات الإضافيين الأخيرين سيعطي $SSR(X_1, X_3 \mid X_2)$.

والسبب في أهمية مجاميع المربعات الإضافية هو أنها تظهر في اختبارات متنوعة حول معاملات الانحدار حيث تكون المسألة المعنيّة هي ما إذا كان يمكن شطب متغيرات مستقلة معينة من نموذج الانحدار. ونتحول فيما يلي إلى مثل هذا الاستخدام لمجاميع المربعات الإضافية.

استخدام مجاميع المربعات الإضافية في اختبار ما إذا كانت β_k بمفردها مساوية للصفر

عندما نرغب في اختبار ما إذا كان يمكن شطب الحد $\beta_k X_k$ من نموذج انحدار متعدد، فالبدائل التي تهمنا هي:

$$H_0 : \beta_k = 0$$

$$H_a : \beta_k \neq 0$$

ونعلم سابقا أن إحصاءة الاختبار (7.46b)

$$t^* = \frac{b_k}{s\{b_k\}}$$

هي الإحصاءة المناسبة لهذا الاختبار.

وبصورة مكافئة، يمكن استخدام أسـلوب الاختبـار الخطّـي العـام الموصـوف في الفقرة (٣-٩). وسنبيّن الآن أن هذا الأسلوب ينطوي على مجاميع مربعـات إضافيـة. لنعتبر نموذج الانحدار من المرتبة الأولى بثلاثة متغيرات مستقلة:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad (8.12)$$ نموذج تام

فلاختبار البدائل:

$$\begin{aligned} H_0 &: \beta_k = 0 \\ H_a &: \beta_k \neq 0 \end{aligned} \qquad (8.13)$$

نقوم بتوفيق النموذج التام ونحصل على مجموع مربعات الخطأ $SSE(F)$. وسـنقوم الآن بعرض ظاهر للمتغيرات في النموذج التام كما يلي:

$$SSE(F) = SSE(X_1 , X_2 , X_3)$$

ودرجات الحرية المرافقة لـ $SSE(F)$ هي $df_F = n - 4$؛ لنتذكر أنه يوجد أربـع معـالم في دالة الانحدار في النموذج التام (8.12).

والنموذج المخفض عندما تكون H_0 في (8.13) صحيحة هو:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i2} + \varepsilon_i \qquad (8.14)$$ نموذج مخفّض

ونقوم الآن بتوفيق هذا النموذج المخفض فنجد:

$$SSE(R) = SSE(X_1 , X_2)$$

وهناك $df_R = n - 3$ درجة حرية تترافق مع هذا النموذج المخفّض.

وإحصاءة الاختبار الخطّي العام (3.69):

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

تصبح هنا:

$$F^* = \frac{SSE(X_1, X_2) - SSE(X_1, X_2, X_3)}{(n-3) - (n-4)} \div \frac{SSE(X_1, X_2, X_3)}{n-4}$$

لاحظ أن الفرق بين مجموعي مربعات الخطأ في البسط هـو مجمـوع المربعـات الإضـافي

:(8.3a)

$$SSE(X_1, X_2) - SSE(X_1, X_2, X_3) = SSR(X_3 \mid X_1, X_2)$$

وبالتالي تكون إحصاءة الاختبار الخطّي العام هنا كما يلي:

$$F^* = \frac{SSR(X_3|X_1 X_2)}{1} \div \frac{SSE(X_1, X_2, X_3)}{n-4} = \frac{MSR(X_3|X_1 X_2)}{MSE(X_1, X_2, X_3)} \quad (8.15)$$

وهكذا نرى أن اختبار ما إذا كانت $\beta_3 = 0$ أم لا هو اختبار هامشي، علما أن X_1 و X_2 كانا من حينهما في النموذج. ونلاحظ أيضا أن لمجموع المربعات الإضافي $SSR(X_3 \mid X_1, X_2)$ درجة واحدة من الحرية تترافق معه، تماما كما أشرنا من قبل.

وتبين إحصاءة الاختبار (8.15) أننا لانحتاج هنا إلى توفيق كل من النموذجين التام والمخفض، كي نستخدم الاختبار الخطّي العام. ويمكن أن تقدم تشغيلة واحدة للحاسب توفيقا للنموذج التام ومجموع المربعات الإضافي المناسب.

مثال. في مثال شحوم الجسم مع المتغيرات المستقلة الثلاثة في الجدول (٨-١) جميعها، نرغب في اختبار ما إذا كان يمكن شطب محيط منتصف الذراع (X_3) من النموذج. وبدائل الاختبار هي تلك المذكورة في (8.13) ويتضمن الجدول (٨-٤) نتائج التحاين من توفيق بالحاسب للنموذج التام (8.12)، متضمنا مجاميع المربعات الإضافية عند إدخال المتغيرات المستقلة بالترتيب X_1 ثم X_2 ثم X_3، وبالتالي تكون إحصاءة الاختبار (8.15) هنا:

$$F^* = \frac{SSR(X_3|X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_3)}{n-4} = \frac{11.45}{1} \div \frac{98.41}{16} = 1.88$$

ومن أجل $\alpha = 0.01$، نحتاج إلى $F(.99;1\ ,\ 16) = 8.53$ وبما أن $F^* = 1.88 \le 8.53$ نستنتج H_0، أي أنه يمكن شطب X_3 من نموذج الانحدار الذي تضمن من حينه X_1 و X_2.

ونلاحظ من الجدول (٨-٢)د أن t^* إحصاءة الاختبار هي هنا:

$$t^* = \frac{b_3}{s\{b_3\}} = \frac{-2.186}{1.596} = -1.37$$

وبما أن $(t^*)^2 = (-1.37)^2 = 1.88 = F^*$ ، نرى أن إحصاءتي الاختبار متكافئتان، وذلك كما في حالة الانحدار الخطّي البسيط تماما.

ملاحظة

تدعى إحصاءة الاختبار F^* في (8.15) لاختبار ما إذا كانت $\beta_3 = 0$ أم لا؛ إحصاءة الاختبار F الجزئي لتمييزها عن الإحصاءة F^* في (7.34b) لاختبار ما إذا كانت جميع المعالم β_k مساوية للصفر، أي لاختبار ما إذا كانت توجد علاقة انحدار بين Y ومجموعة المتغيرات المستقلة. وهذا الاختبار الأخير يدعى اختبار F الإجمالي.

استخدام مجاميع المربعات الإضافية لاختبار ما إذا كانت عدّة معالم β_k مساوية للصفر

كثيرا ما نهتم في الانحدار المتعدد بما إذا كان يمكن شطب عدّة حدود في نموذج الانحدار. وعلى سبيل المثال، قد نرغب في معرفة ماإذا كان يمكن شطب $\beta_2 X_2$ و$\beta_3 X_3$ من النموذج التام (8.12) والبدائل هنا هي:

$$H_0 : \beta_2 = \beta_3 = 0 \tag{8.16}$$

H_a: ليس كل من β_3 و β_2 مساويا للصفر

ووفقا لأسلوب الاختبار الخطي العام، يكون النموذج المخفض تحت الفرضية H_0

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \tag{8.17}$$ نموذج مخفض

ومجموع مربعات الخطأ للنموذج المخفض هو:

$$SSE(R) = SSE(X_1)$$

ويترافق مع مجموع مربعات الخطأ هذا $df_R = n - 2$ درجة من الحرية.

وهكذا تصبح إحصاءة الاختبار (3.69) هنا:

$$F^* = \frac{SSE(X_1) - SSE(X_1, X_2, X_3)}{(n-2)-(n-4)} \div \frac{SSE(X_1, X_2, X_3)}{n-4}$$

ومرة ثانية نجد أن الفرق بين مجموعي مربعات الخطأ في البسط هو مجموع مربعات إضافي، ونعني:

$$SSE(X_1) - SSE(X_1, X_2, X_3) = SSR(X_2, X_3 \mid X_1)$$

وبالتالي، تصبح إحصاءة الاختبار:

$$F^* = \frac{SSR(X_2, X_3 \mid X_1)}{2} \div \frac{SSE(X_1, X_2, X_3)}{n-4} = \frac{MSR(X_2, X_3 \mid X_1)}{MSE(X_1, X_2, X_3)} \tag{8.18}$$

ونلاحظ أنه يترافق مع $SSR(X_2, X_3 \mid X_1)$ درجتان من الحرية، وذلك كما أشرنا سابقا.

مثال. في مثال شحوم الجسم بثلاثة متغيرات مستقلة، نرغب في اختبار ما إذا كان يمكن شطب كل من محيط الفخذ (X_2) ومحيط منتصف الذراع (X_3) من نموذج الانحدار التام (8.12). والبدائل هي تلك المذكورة في (8.16). ويمكن الحصول على مجاميع المربعات المناسبة من الجدول (٨-٤)، مستخدمين (8.11):

$$SSR(X_2, X_3 \mid X_1) = SSR(X_2 \mid X_1) + SSR(X_3 \mid X_1, X_2) = 33.17 + 11.54 = 44.71$$

وبالتالي تكون إحصاءة الاختبار (8:18):

$$F^* = \frac{SSR(X_2, X_3 | X_1)}{2} \div MSE(X_1, X_2, X_3) = \frac{44.71}{2} \div 6.15 = 3.63$$

ومن أجل $\alpha = 0.01$ نحتاج إلى $F(.99; 2, 16) = 6.23$. وبما أن $F^* = 3.63 \leq 6.32$ فنستنتج H_0، أي أنه يمكن شطب X_2 و X_3 من نموذج الانحدار الذي تضمن من حينه X_1.

ملاحظة

لاختبار ما إذا كانت β_k بمفردها مساوية للصفر تتوافر لنا إحصاءتا اختبار متكافئتان: إحصاءة الاختبار t^* في (7.46b) و F^* إحصاءة الاختبار الخطّي العام في (3.69). وعند اختبار ما إذا كانت عدة معالم β_k مساوية للصفر لايتوافر لنا إلا إحصاءة الاختبار الخطّي العام F^*.

(٨-٢) اختبار فرضيات تتعلق بمعاملات الانحدار في انحدار متعدد

ناقشنا سابقا كيفية القيام بعدّة أنواع من الاختبارات المتعلقة بمعاملات الانحدار في نموذج انحدار متعدّد. ولتمام المناقشة نلخّص هذه الاختبارات هنا ثم نتابع دراسة أنواع إضافية من الاختبارات.

اختبار ما إذا كانت جميع المعاملات β_k مساوية للصفر

هذا هو اختبار F الإجمالي (7.34) ما إذا كانت توجد علاقة انحدار أم لا بين المتغير التابع Y ومجموعة المتغيرات المستقلة. والبدائل هي:

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0 \tag{8.19}$$

H_a: ليست جميع المعاملات β_k $(k = 1, \ldots, p-1)$ مساوية للصفر

وإحصاءة الاختبار هي:

$$F^* = \frac{SSR(X_1,\ldots,X_{p-1})}{p-1} \div \frac{SSE(X_1,\ldots,X_{p-1})}{n-p} = \frac{MSR}{MSE}$$

وإذا كانت H_0 صحيحة فإن $F^* \sim F(p-1, n-p)$ وتؤدي القيم الكبيرة لـ F^* إلى استنتاج H_a.

اختبار ما إذا كانت معلمة بمفردها β_k مساوية للصفر

هذا هو اختبار F الجزئي لما إذا كانت معلمة الانحدار β_k بالذات مساوية للصفر. والبدائل هي:

$$\begin{aligned} H_0 &: \beta_k = 0 \\ H_a &: \beta_k \neq 0 \end{aligned} \tag{8.21}$$

وإحصاءة الاختبار هي:

$$\begin{aligned} F^* &= \frac{SSR(X_k \mid X_1,\ldots,X_{k-1},X_{k+1},\ldots,X_{p-1})}{1} \div \frac{SSE(X_1,\ldots,X_{p-1})}{n-p} \\ &= \frac{MSR(X_k \mid X_1,\ldots,X_{k-1},X_{k+1},\ldots,X_{p-1})}{MSE} \end{aligned} \tag{8.22}$$

وإذا كانت H_0 صحيحة فإن $F^* \sim F(1, n-P)$. وتقود القيم الكبيرة لـ F^* إلى استنتاج H_a. وتسمح لنا حزم الحاسب التي تقدم مجاميع المربعات الإضافية باستخدام هذا الاختبار دون الإضطرار إلى توفيق النموذج المخفض.

وإحصاءة الاختبار المكافئة ، كما رأينا ، هي (7.46b) :

$$t^* = \frac{b_k}{s\{b_k\}} \tag{8.23}$$

وإذا كانت H_0 صحيحة فإن $t^* \sim t(n-P)$. وتقود القيم الكبيرة لـ $|t^*|$ إلى استنتاج H_a.

وبما أن الاختبارين متكافئان ، فالاختيار يتم عادة وفقا للمعلومات المتوافرة التي تقدمها مطبوعة مُخرجات حزمة الحاسب.

اختبار ما إذا كانت بعض المعالم β_k مساوية للصفر

وهذا اختبار جزئي آخر. والبدائل هي:

$$H_0 : \beta_q = \beta_{q+1} = \ldots = \beta_{p-1} = 0 \tag{8.24}$$

H_a: ليست جميع β_k المذكورة في H_0 مساوية للصفر

ولاعتبارات السهولة، نرتب النموذج بحيث تكون المعاملات الـ $p - q$ الأخيرة هي تلك التي نريد اختبارها. وإحصاءة الاختبار هي:

$$F^* = \frac{SSR(X_q \ldots, X_{p-1} \mid X_1, \ldots, X_{q-1})}{p-q} \div \frac{SSE(X_1, \ldots, X_{p-1})}{n-p} \tag{8.25}$$
$$= \frac{MSR(X_q \ldots, X_{p-1} \mid X_1, \ldots, X_{q-1})}{MSE}$$

وإذا كانت H_0 صحيحة فإن $F^* \sim F(P - q, n - P)$. وتؤدي القيم الكبيرة لـ F^* إلى استنتاج H_a.

ونلاحظ أن إحصاءة الاختبار (8.25) تحيط بالحالتين السابقتين. فإذا كان $q = 1$، فإن الاختبار هو ما إذا كانت جميع معاملات الانحدار مساوية للصفر. وإذا كان $q = P - 1$ فإن الاختبار هو ما إذا كانت معلمة انحدار بمفردها مساوية للصفر. ونلاحظ أيضا أنه يمكن حساب إحصاءة الاختبار (8.25) دون الاضطرار إلى توفيق النموذج المخفض إذا كان برنامج الحاسب يقدم مانحتاجه من مجاميع المربعات الإضافية.

$$SSR(X_q, \ldots, X_{p-1} \mid X_1, \ldots, X_{q-1}) = SSR(X_q \mid X_1, \ldots, X_{q-1}) + \ldots + SSR(X_{p-1} \mid X_1, \ldots, X_{p-2}) \tag{8.26}$$

اختبارات أخرى

عندما نرغب في اختبارات حول معاملات الانحدار لاتنطوي على اختبار ما إذا كانت معلمة أو عدة معالم تساوي الصفر، فلا يمكن استخدام مجاميع المربعات الإضافية. ويتطلب الاختبار الخطّي العام توفيقين منفصلين للنموذجين التام والمخفض وعلى سبيل المثال، في نموذج تام يتضمن ثلاثة متغيرات X :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad \text{نموذج تام} \tag{8.27}$$

قد نرغب في اختبار:

$$H_0: \beta_1 = \beta_2$$
$$H_a: \beta_1 \neq \beta_2 \tag{8.28}$$

والطريقة هي أن نقوم بتوفيق النموذج التام (8.27) ثم النموذج المخفض:

$$Y_i = \beta_0 + \beta_c(X_{i1} + X_{i2}) + \beta_3 X_{i3} + \varepsilon_i \quad (8.29)$$ نموذج مخفض

حيث ترمز β_c للمعامل المشترك لـ β_1 و β_2 تحت H_0 وبمتغير X جديد يقابله هو $X_{i1}+X_{i2}$. وعندئذ نستخدم إحصاءة الاختبار العامة F^* كما وردت في (3.69) بدرجة واحدة و $n - 4$ درجة من الحرية.

وكمثال آخر، لايمكن فيه استخدام مجاميع المربعات الإضافية، نذكر الاختبار التالي في نموذج الانحدار (8.27):

$$H_0 : \beta_1 = 3, \beta_3 = 5 \quad (8.30)$$

H_a : ليست المعادلتان المذكورتان في H_0 صحيحتين معا

وسيكون النموذج المخفض هنا:

$$Y_i - 3X_{i1} - 5X_{i1} = \beta_0 + \beta_2 X_{i2} + \varepsilon_i \quad (8.31)$$ نموذج مخفض

لاحظ أن المتغير التابع الجديد في النموذج المخفض هو $Y - 3X_1 - 5X_3$ باعتبار $\beta_1 X_1$ و$\beta_3 X_3$ ثوابت معروفة تحت H_0. ونستخدم عندئذ إحصاءة الاختبار الخطي العام F^* في (3.69) بدرجتين و $n - 4$ درجة من الحرية.

(٨-٣) معاملات التحديد الجزئية

لاتفيدنا مجاميع المربعات الإضافية فقط في اختبارات حول معاملات الانحدار في نموذج انحدار متعدد، ولكنها تواجهنا أيضا في مقاييس تدعى معاملات التحديد الجزئية، وهي مقاييس لدرجة تواجد صلة أو علاقة. وكما نذكر، فإن معامل التحديد المتعدد R^2 يقيس التخفيض النسبي في تغير Y الذي حققه إدخال مجموعة المتغيرات X المعنية بكاملها في النموذج. وفي المقابل، فإن **معامل التحديد الجزئي** يقيس المساهمة الهامشية لمتغير واحد من المتغيرات X ، عندما كانت جميع المتغيرات الأخرى مشمولة من حينها في النموذج.

متغيران مستقلان

لنعتبر نموذج انحدار متعدد من المرتبة الأولى بمتغيرين مستقلين، كما أُعطي في (7.1).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

فيقيس $SSE(X_2)$ التغير في Y عندما يكون X_2 مشمولا في النموذج بمفرده. ويقيس $SSE(X_1, X_2)$ التغير في Y عندما يكون X_1 وX_2 مشمولين في النموذج. وبالتالي فإن التخفيض الهامشي النسبي في تغير Y المترافق مع X_1 عندما كان X_2 من حينه في النموذج هو:

$$\frac{SSE(X_2)-SSE(X_1,X_2)}{SSE(X_2)}=\frac{SSR(X_1|X_2)}{SSE(X_2)}$$

وهذا القياس هو معامل التحديد الجزئي بين Y و X_1 علما أن X_2 في النموذج. ونرمز لهذا القياس بـ $r^2_{Y1.2}$:

$$r^2_{Y1.2}=\frac{SSE(X_2)-SSE(X_1,X_2)}{SSE(X_2)}=\frac{SSR(X_1|X_2)}{SSE(X_2)} \tag{8.32}$$

وهكذا يقيس $r^2_{Y1.2}$ التخفيض النسبي في تغير Y المتبقي بعد أن كان X_2 في النموذج والذي يتمُّ اكتسابه بضم X_1 أيضا إلى النموذج.

ونعرف معامل التحديد الجزئي بين Y وX_2 ، علما أن X_1 موجود في النموذج، كما يلي:

$$r^2_{Y2.1}=\frac{SSR(X_2|X_1)}{SSE(X_1)} \tag{8.33}$$

حالة عامة

ويمكن التعميم مباشرة إلى معاملات تحديد جزئي لثلاثة متغيرات مستقلة في النموذج أو أكثر. وعلى سبيل المثال:

$$r^2_{Y1.23} =\frac{SSR(X_1|X_2,X_3)}{SSE(X_2,X_3)} \tag{8.34}$$

$$r^2_{Y2.13} =\frac{SSR(X_2|X_1,X_3)}{SSE(X_1,X_3)} \tag{8.35}$$

$$r^2_{Y3.12} = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)} \quad (8.36)$$

$$r^2_{Y4.123} = \frac{SSR(X_4|X_1, X_2, X_3)}{SSE(X_1, X_2, X_3)} \quad (8.37)$$

في أدلة r^2 لاحظ أن مايدخل منها على يسار النقطة يبين على التوالي المتغير المعتبر كاستجابة، والمتغير الذي أضيف. ومايدخل منها على يمين النقطة يبين المتغيرات X التي كانت موجودة من حينها في النموذج.

مثال

في مثال شحوم الجسم، يمكن الحصول على تشكيلة من معاملات التحديد الجزئي. وفيما يلي ثلاثة منها (جدول ٨-٢ و ٨-٤):

$$r^2_{Y2.1} = \frac{SSR(X_2|X_1)}{SSE(X_1)} = \frac{33.17}{143.12} = 0.232$$

$$r^2_{Y3.12} = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)} = \frac{11.54}{109.95} = 0.105$$

$$r^2_{Y1.2} = \frac{SSR(X_1|X_2)}{SSE(X_2)} = \frac{3.47}{113.42} = 0.031$$

وهكذا فإن $SSE(X_1)$ قد انخفض، بنسبة 23.2 بالمائة عند إضافة X_2 إلى نموذج الانحدار المتضمن لِ X_1. ومجموع مربعات الخطأ للنموذج المتضمن لـ X_1 و X_2 قد انخفض أيضا بنسبة 10.5 في المائة فقط عند إضافة X_3 إلى النموذج. وأخيرا ، إذا كان نموذج الانحدار قد تضمن من حينه X_2 فإن إضافة X_1 فقط ستخفض $SSE(X_2)$ بنسبة 3.1 بالمائة.

تعليقات

١- يمكن أن تتخذ معاملات التحديد الجزئي قيما بين 0 و 1 كما يشير التعريف مباشرة.

٢- يمكن تفسير معامل تحديد جزئي كمعامل تحديد بسيط. فلنعتبر نموذج انحدار متعدد بمتغيرين مستقلين. ولنفرض أننا حدرنا Y على X_2 وحصلنا على الرواسب:

$$Y_i - \hat{Y}_i(X_2)$$

حيث ترمز $\hat{Y}_i(X_2)$ للقيم التوفيقية لِـ Y عندما يكون X_2 في النموذج.

ولنفترض أننا حدرنا أيضا X_1 على X_2 وحصلنا على الرواسب:

$$X_{i1} - \hat{X}_{i1}(X_2)$$

حيث ترمز $\hat{X}_{i1}(X_2)$ للقيم التوفيقية لـ X_1 في انحدار X_1 على X_2. فمعامل التحديد البسيط r^2 بين هاتين المجموعتين من الرواسب يساوي معامل التحديد الجزئي $r^2_{Y1.2}$. وهكذا يقيس هذا المعامل العلاقة بين Y و X_1 عند تعديل هذين المتغيرين كليهما من أجل علاقاتهما الخطّية بـ X_2.

معاملات الارتباط الجزئي

يدعى الجذر التربيعي لمعامل التحديد الجزئي معامل الارتباط الجزئي. ويُعطى نفس إشارة معامل الانحدار المقابل في دالة الانحدار التوفيقية، وكثيرا مانستخدم معاملات الارتباط الجزئي في التطبيقات العملية، مع أنها لاتمتلك معنى واضحا كوضوح معاملات التحديد الجزئية.

ولدينا في مثال شحوم الجسم:

$$r_{Y2.1} = \sqrt{0.232} = 0.482$$
$$r_{Y3.12} = -\sqrt{0.105} = -0.324$$
$$r_{Y1.2} = \sqrt{0.031} = 0.176$$

لاحظ أن المعاملين $r_{Y2.1}$ و $r_{Y1.2}$ موجبان لأن b_2=0.6594 و 0.2224= b_1 كما نرى من الجدول (٨-٢)جـ وبصورة مماثلة، $r_{Y3.12}$ سالب لأن 2.186- = b_3 كما نرى من الجدول (٨-٢)د.

وكثيرا مانستخدم معاملات الارتباط الجزئي في روتينيات الحاسب بغية إيجاد المتغير المستقل الأفضل الذي سنختاره في الخطوة التالية لضمه إلى نموذج الانحدار، وسنناقش مثل هذا الاستخدام في الفصل الثاني عشر.

ملاحظة

يمكن التعبير عن معاملات التحديد الجزئي بدلالة معاملات الارتباط البسيطة أو

معاملات الارتباط الجزئي الأخرى. فمثلا:

$$r^2_{Y2.1} = \frac{(r_{Y2} - r_{12} r_{Y1})^2}{(1-r^2_{12})(1-r^2_{Y1})} \tag{8.38}$$

$$r^2_{Y2.13} = \frac{(r_{Y2.3} - r_{12.3} r_{Y1.3})^2}{(1-r^2_{12.3})(1-r^2_{Y1.3})} \tag{8.39}$$

حيث يرمز r_{Y1} لمعامل الارتباط البسيط بين Y و X_1 ويرمز r_{12} لمعامل الارتباط البسيط بين X_1 و X_2 وهكذا. والتعميمات مباشرة ولاصعوبة فيها.

(٨–٤) نموذج انحدار متعدّد معياري

يُستخدم الشكل المعياري لنموذج الانحدار المتعدّد العام (7.7) للتحكم بأخطاء تدوير الأرقام العشرية في حسابات المربعات الدنيا، ولكي نتمكن من القيام بمقارنات بين معاملات الانحدار المقدّرة بوحدات قياس مشتركة.

أخطاء تدوير الأرقام العشرية في حسابات المربعات الدنيا

يمكن أن تكون نتائج المربعات الدنيا حساسة لتدوير أرقام عشرية في البيانات وذلك في مراحل متوسطة من الحسابات، وعندما يكون عدد المتغيرات المستقلة صغيرا ـ لنقل ثلاثة أو أقل ـ يمكن التحكم بتأثيرات التدوير عن طريق حَمْل عدد كاف من الأرقام العشرية في المراحل المتوسطة من الحسابات. وفي الحقيقة، تستخدم معظم برامج الانحدار الحاسوبية دقة حسابية مضاعفة (مثلا، استخدام 16 رقما عشريا بدلا من 8 أرقام عشرية) في جميع الحسابات كي تتحكم في تأثيرات التدوير. ومع عدد كبير من المتغيرات X لاتزال هناك إمكانية لظهور تأثيرات جدية للتدوير بالرغم من استخدام العديد من الأرقام العشرية في المراحل المتوسطة من الحسابات.

وتميل أخطاء التدوير إلى الدخول في حسابات المربعات الدنيا عند حساب معكوس $\mathbf{X'X}$ في المقام الأول. وبالطبع يمكن أن تتضخم أية أخطاء في $(\mathbf{X'X})^{-1}$ عند حساب $\mathbf{b}$ وإحصاءات لاحقة أخرى. وتكون خطورة أخطاء تدوير جدية في $(\mathbf{X'X})^{-1}$ خطورة عظيمة، على وجه الخصوص، عندما : (١) يكون لـ $\mathbf{X'X}$ محددا قريبا من

الصفر و/أو (٢) اختلاف شديد في مقادير عناصر **X'X**. ويبرز الشرط الأول عندما تكون جميع المتغيرات المستقلة أو بعضها مرتبطة فيما بينها. وسنناقش هذه الحالة في الفقرة (٨-٥).

ويبرز الشرط الثاني عندما يكون للمتغيرات مقادير مختلفة اختلافا شديدا بحيث تغطي عناصر المصفوفة **X'X** مدى واسعا من الأعداد، مثلا من 15 إلى 49,000,000. والحل في مثل هذا الشرط هو تحويل المتغيرات وبالتالي إعادة صياغة معالم نموذج الانحدار.

والتحويل الذي سنتبناه يدعى تحويل الارتباط. وهو يجعل جميع المقادير في المصفوفة **X'X** محسوبة بدلالة المتغيرات الجديدة، واقعة بين 1- و 1+، بما فيه الطرفان، وهكذا تصبح حسابات معكوس المصفوفة من حيث خضوعها لأخطاء التدوير العائدة إلى مفارقات كبيرة في المقادير أقل بكثير مما كانت عليه في المتغيرات الأصلية. وكثير من حزم الحاسب الخاصة بالانحدار تستخدم بصورة آلية هذا التحويل للحصول على نتائج الانحدار الأساسية ثم تعيد تحويل هذه النتائج بدلالة المتغيرات الأصلية.

نقص قابلية المقارنة في معاملات الانحدار

والصعوبة الثانية في نموذج الانحدار المتعدد غير المعياري (7.7) هو أن معاملات الانحدار غير قابلة، في المعتاد، للمقارنة، بسبب الفروق في وحدات القياس المستخدمة. ونذكر مثالين:

١- عند اعتبار دالة الاستجابة التوفيقية:

$$\hat{Y}=200+20{,}000X_1+0.2X_2$$

قد يميل المرء لاستنتاج أن المتغير المستقل المهم الوحيد هو X_1 وأن لـ X_2 تأثيرا طفيفا على المتغير التابع Y. وقليل من التأمل ينبغي أن يجعل المرء حذرا من هذه النتيجة. ذلك لأننا لانعلم الوحدات التي نقيس بها. فلنفرض أن الوحدات هي:

Y بالدولار.

X_1 بآلاف الدولارات.

X_2 بالسنتات.

ففي هذه الحالة سيكون تأثير زيادة قدرها 1000 $ في X_1 على متغير الاستجابة مع بقاء X_2 ثابتا، هو بالضبط نفس تأثير زيادة قدرها 1000 $ في X_2 مع بقاء X_1 ثابتا، وذلك بالرغم من الفرق في معاملات الانحدار.

٢ـ في مثال شركة زارثان في الجدول (٧-٢) لا يمكن القيام بأية مقارنة بين b_1 و b_2 لأن b_1 تقاس بوحدة هي كروز لكل 1000 شخص، بينما تقاس b_2 بوحدة هي كروز لكل دولار من الدخل الفردي المصرح به.

تحويل الارتباط

يساعد استخدام تحويل الارتباط في التحكم بأخطاء التدوير ويجعل وحدات معاملات الانحدار قابلة للمقارنة. وسنصِف أولا تحويل الارتباط ثم نموذج الانحدار المعياري الناتج.

وتحويل الارتباط هو تعديل بسيط للمعايرة المعتادة لمتغير. إذ تنطوي معايرة متغير، كما في (1.34)، على أخذ الفرق بين كل مشاهدة ومتوسط جميع المشاهدات. ثم التعبير عن هذه الفروق بوحدة قياس هي الانحراف المعياري للمشاهدات. وهكذا تكون المعايرات المعتادة للمتغير التابع Y ، والمتغيرات المستقلة X_1،...، X_{p-1} كما يلي:

$$\frac{Y_i-\overline{Y}}{s_Y} \qquad (8.40a)$$

$$\frac{X_{ik}-\overline{X}_k}{s_k} \qquad (k=1,\ldots,p\text{-}1) \qquad (8.40b)$$

حيث $\overline{Y}$ و $\overline{X}_k$ هما متوسطا Yو X_k على الترتيب، و s_Y و s_k هما، على الترتيب، الانحرافان المعياريان المعرّفان كما يلي:

$$s_Y=\sqrt{\frac{\sum_i (Y_i-\overline{Y})^2}{n-1}} \qquad (8.40c)$$

$$s_k=\sqrt{\frac{\sum_i (X_{ik}-\overline{X}_k)^2}{n-1}} \qquad (k=1,\ldots,p\text{-}1) \qquad (8.40d)$$

ويستخدم تحويل الارتباط الدالة التالية في المتغيرات المعيارية في (8.40):

$$Y_i' = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \overline{Y}}{s_Y}\right) \tag{8.41a}$$

$$X_{ik}' = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ik} - \overline{X}_k}{s_k}\right) \quad (k = 1,\ldots,p\text{-}1) \tag{8.41b}$$

نموذج انحدار معياري

يدعى نموذج الانحدار في المتغيرات Y' و X_k' كما عرفناها في تحويل الارتباط في (8.41)، نموذج الانحدار المعياري، وهو كما يلي:

$$Y_i' = \beta_1' X_{i1}' + \ldots + \beta_{p-1}' X_{i,p-1}' + \varepsilon_i' \tag{8.42}$$

وسبب عدم وجود معلمة الجزء المقطوع في نموذج الانحدار المعياري (8.42) هو أن حسابات المربعات الدنيا ستقود دائما إلى حد للجزء المقطوع يساوي الصفر، هذا إذا وضعنا معلمة جزء مقطوع في النموذج.

ومن السهل تبيان أن العلاقات بين المعالم الجديدة β_1'،...، β_{p-1}' والمعالم الأصلية β_0، β_1،...،β_{p-1} في نموذج الانحدار المتعدد المعتاد *(7.7)* هي كما يلي:

$$\beta_k = \left(\frac{s_Y}{s_k}\right)\beta_k' \qquad (k = 1,\ldots,p\text{-}1) \tag{8.43a}$$

$$\beta_0 = \overline{Y} - \beta_1\overline{X}_1 - \ldots - \beta_{p-1}\overline{X}_{p-1} \tag{8.43b}$$

وهكذا فإن معاملات الانحدار الجديدة β_k' ومعاملات الانحدار الأصلية β_k $(K = 1,\ldots,P - 1)$ تعلق ببعضها من خلال عوامل سلّمية بسيطة تنطوي على نسب انحرافات معيارية.

المصفوفة X'X بدلالة المتغيرات بعد التحويل

كي نكون قادرين على دراسة الطبيعة الخاصة للمصفوفة **X'X** والمعادلات الناظمية للمربعات الدنيا بعد تحويل المتغيرات وفقا لتحويل الارتباط، سنحتاج إلى تعريف مصفوفتين تتضمنان معاملات ارتباط بسيط. وتدعى المصفوفة الأولى، ونرمز لها بـ $\mathbf{r}_{XX}$ مصفوفة الارتباط للمتغيرات X وعناصرها معاملات الارتباط البسيط بين كافة الأزواج من المتغيرات X وتُعرف هذه المصفوفة كما يلي:

$$\underset{(p-1)\times(p-1)}{\mathbf{r}_{XX}} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1,p-1} \\ r_{21} & r_1 & \cdots & r_{2,p-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ r_{p-1,1} & r_{p-1,2} & \cdots & 1 \end{bmatrix} \tag{8.44}$$

ويرمز r_{12} هنا، لمعامل الارتباط البسيط بين X_1 و X_2 ويرمز r_{13} لمعامل الارتباط البسيط بين X_1 و X_2، وهكذا. لاحظ أن القطر الرئيس يتألف من الأعداد 1 لأن معامل الارتباط البسيط بين المتغير ونفسه هو الواحد. ومصفوفة الارتباط $\mathbf{r}_{XX}$ متناظر. لنتذكر أن $r_{kk'} = r_{k'k}$ وبسبب تناظر هذه المصفوفة ستحذف مُخرجات الحاسب، الغالب، القطاع المثلث الأدنى أو الأعلى من العناصر.

والمصفوفة الثانية، التي سنرمز لها بـ r_{YX} هي متجه يتضمن معاملات الارتباط البسيط بين المتغير التابع Y وكل من المتغيرات X ، ونرمز لها بـ r_{Y1} ، r_{Y2} الخ:

$$\underset{(p-1)\times 1}{\mathbf{r}_{YX}} = \begin{bmatrix} r_{Y1} \\ r_{Y1} \\ \cdot \\ \cdot \\ \cdot \\ r_{Y,p-1} \end{bmatrix} \tag{8.45}$$

ونحن جاهزون الآن لاعتبار المصفوفة $\mathbf{X'X}$ للمتغيرات بعد التحويل في نموذج الانحدار المعياري (8.42) والمصفوفة $\mathbf{X}$ هنا هي:

$$\underset{n\times(p-1)}{\mathbf{X}} = \begin{bmatrix} X'_{11} & \cdots & X'_{1,p-1} \\ X'_{21} & \cdots & X'_{2,p-1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ X'_{n1} & \cdots & X'_{n,p-1} \end{bmatrix} \tag{8.46}$$

لنتذكر أن نموذج الانحدار المعياري (8.42) لايتضمن حد الجزء المقطوع وبالتالي لايوجد عمود من الأعداد 1 في المصفوفة $\mathbf{X}$. ويمكن تبيان أن المصفوفة $\mathbf{X'X}$

للمتغيرات بعد التحويل هي ببساطة مصفوفة الارتباط للمتغيرات X المعرفة في (4ا

$$\underset{(p-1)\times(p-1)}{\mathbf{X'X}} = \mathbf{r}_{XX} \tag{8.47}$$

وبما أن المصفوفة $\mathbf{X'X}$ للمتغيرات بعد التحويل تتألف من معاملات الارتباط المتغيرات X ، فجميع عناصرها تقع بين 1- و 1+ وتكون، هكذا، من الدرجة نة في الكبر. وكما أشرنا سابقا، فإن هذا يمكن أن يشكل عونا عظيما في السيطرة أخطاء التدوير عند عكس المصفوفة $\mathbf{X'X}$.

ملاحظة

نوضّح كون المصفوفة $\mathbf{X'X}$ للمتغيرات الجديدة هي مصفوفة الارتباط نفسها للمت X بحسابنا لعنصرين من المصفوفة:

١- ففي الزاوية اليسرى العليا من $\mathbf{X'X}$ لدينا:

$$\sum (X'_{i1})^2 = \sum\left(\frac{X_{i1}-\overline{X}_1}{\sqrt{n-1}\,s_1}\right)^2 = \frac{\Sigma(X_{i1}-\overline{X}_1)^2}{n-1} \div s_1^2 = 1$$

٢- وفي السطر الأول والعمود الثاني من $\mathbf{X'X}$، لدينا:

$$X'_{i1}X'_{i2} = \sum\left(\frac{X_{i1}-\overline{X}_1}{\sqrt{n-1}\,s_1}\right)\left(\frac{X_{i2}-\overline{X}_2}{\sqrt{n-1}\,s_2}\right)$$

$$= \frac{1}{n-1}\,\frac{\Sigma(X_{i1}-\overline{X}_1)(X_{i2}-\overline{X}_2)}{s_1 s_2} = \frac{\Sigma(X_{i1}-\overline{X}_1)(X_{i2}-\overline{X}_2)}{\left[\Sigma(X_{i1}-\overline{X}_1)^2\,\Sigma(X_{i2}-\overline{X}_2)^2\right]^{1/2}}$$

ولكن هذا يساوي r_{12} ، معامل الارتباط بين X_1 و X_2 وفقا لـ (3.75).

معاملات الانحدار المعيارية المقدّرة

معادلات المربعات الدنيا الناظمية (7.20) لنموذج الانحدار المتعدد العادي ه

$$\mathbf{X'X\,b} = \mathbf{X'Y}$$

ويمكن التعبير عن مقدرات المربعات الدنيا (7.21):

$$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

ببساطة أكثر بالنسبة للمتغيرات بعد التحويل. من السهل تبيان أن $\mathbf{X'Y}$ تصبح أجل المتغيرات بعد التحويل:

$$\underset{(p-1)\times 1}{\mathbf{X'Y}} = \mathbf{r}_{YX} \tag{8.48}$$

حيث $\mathbf{r}_{YX}$ معرفة في (8.45) كمتجه معاملات الارتباط البسيط بين Y وكل من المتغيرات X. ومن (8.47) و(8.48) نجد الآن أن معادلات المربعات الدنيا الناظمية ومقدرات معاملات الانحدار لنموذج الانحدار المعياري (8.42) هي كما يلي:

$$\mathbf{r}_{XX}\mathbf{b} = \mathbf{r}_{YX} \tag{8.49a}$$

$$\mathbf{b} = \mathbf{r}_{XX}^{-1}\mathbf{r}_{YX} \tag{8.49b}$$

حيث:

$$\underset{(p-1)\times 1}{\mathbf{b}} = \begin{bmatrix} b'_1 \\ b'_2 \\ \cdot \\ \cdot \\ \cdot \\ b'_{p-1} \end{bmatrix} \tag{8.49c}$$

وفي الغالب تدعى معاملات الارتباط b'_1،...، b'_{p-1} معاملات الانحدار المعيارية.

ونقوم بالعودة إلى معاملات الانحدار المقدرة لنموذج الانحدار (7.7) في *المتغيرات الأصلية* باستخدام العلاقات:

$$b_k = \left(\frac{s_Y}{s_k}\right)b'_k \qquad (k = 1,\ldots,p\text{-}1) \tag{8.50a}$$

$$b_0 = \overline{Y} - b_1\overline{X}_1 - \ \ldots\ - b_{p-1}\overline{X}_{p-1} \tag{8.50b}$$

ملاحظة

عندما يكون عدد المتغيرات X في نموذج الانحدار $p\text{-}1 = 2$ يمكننا أن نرى بسهولة الصيغة الجبرية لمعاملات الانحدار المعيارية، ولدينا:

$$\mathbf{r}_{XX} = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \tag{8.51a}$$

$$\mathbf{r}_{YX} = \begin{bmatrix} r_{Y1} \\ r_{Y2} \end{bmatrix} \tag{8.51b}$$

$$\mathbf{r}_{XX}^{-1} = \frac{1}{1-r_{12}^2}\begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \tag{8.51c}$$

وبالتالي نحصل وفقا لـ (8.49d) على:

$$\mathbf{b}=\frac{1}{1-r_{12}^2}\begin{bmatrix}1 & -r_{12}\\ -r_{12} & 1\end{bmatrix}\begin{bmatrix}r_{Y1}\\ r_{Y2}\end{bmatrix}=\frac{1}{1-r_{12}^2}\begin{bmatrix}r_{Y1}-r_{12}r_{Y2}\\ r_{Y2}-r_{12}r_{Y1}\end{bmatrix} \quad (8.52)$$

وهكذا:

$$b_1'=\frac{r_{Y1}-r_{12}r_{Y2}}{1-r_{12}^2} \quad (8.52a)$$

$$b_2'=\frac{r_{Y2}-r_{12}r_{Y1}}{1-r_{12}^2} \quad (8.52b)$$

جدول (٨ -٥) تحويل الارتباط لبيانات شركة زارثان

(أ) البيانات الأصلية

دخل الفرد المصرح به X_{i2}	مجتمع الهدف X_{i1}	المبيعات Y_i	المنطقة i
2,450	274	162	1
3,254	180	120	2
.	.	.	.
.	.	.	.
.	.	.	.
2,605	370	212	15
$\bar{X}_2=2,961.9$	$\bar{X}_1=241.73$	$\bar{Y}=150.60$	
$s_2=730.64$	$s_1=116.83$	$s_Y=62.049$	

(ب) البيانات بعد التحويل

X_{i2}'	X_{i1}'	Y_i'	i
−0.18724	0.07381	0.04910	1
0.10686	−0.14122	−0.13180	2
.	.	.	.
.	.	.	.
.	.	.	.
−0.13054	0.29342	0.26447	15

(جـ) النموذج التوفيقي المعياري

$$\hat{Y}'=0.9339X_1'+0.1083X_2'$$

مثال

يعيد الجدول (٨ـ٥)أ (جزئيا) البيانات الأصلية لمثال شركة زارثان (جدول ٧ـ٢) ويتضمن الجدول (٨ـ٥)ب البيانات بعد تحويلها وفقا لتحويل الارتباط (8.41) ولتوضيح حسابات البيانات بعد التحويل، نجد، مستخدمين المتوسطات والانحرافار

المعيارية في الجدول (٨ـ٥)أ:

$$Y_1'=\frac{1}{\sqrt{n-1}}\left(\frac{Y_1-\overline{Y}}{s_Y}\right)=\frac{1}{\sqrt{15-1}}\left(\frac{162-150.60}{62.049}\right)=0.04910$$

$$X_{11}'=\frac{1}{\sqrt{n-1}}\left(\frac{X_{11}-\overline{X}_1}{s_1}\right)=\frac{1}{\sqrt{15-1}}\left(\frac{274-241.73}{116.83}\right)=0.07381$$

$$X_{12}'=\frac{1}{\sqrt{n-1}}\left(\frac{X_{12}-\overline{X}_2}{s_2}\right)=\frac{1}{\sqrt{15-1}}\left(\frac{2,450-2,961.9}{730.64}\right)=-0.18724$$

وعند توفيق نموذج الانحدار المعياري (8.42) للبيانات بعد التحويل باستخدام حزمة انحدار حاسوبية، نحصل على النموذج التوفيقي في الجدول (٨-٥)جـ:

$$\hat{Y}'=0.9339X_1'+0.1083X_2'$$

وهكذا فإن زيادة بمقدار انحراف معياري واحد في X_1 (المجتمع الهدف)، مع تثبيت X_2، يؤدي إلى زيادة في توقع المبيعات (بوحدة هي الانحراف المعياري لِ Y) أكبر بكثير من الزيادة التي يؤدي إليها زيادة انحراف معياري واحد في X_2 (المدخل الفردي المصرح به) مع تثبيت X_1.

وللعودة من معاملات الانحدار المعيارية b_1' و b_2' إلى معاملات انحدار النموذج بمتغيراته الأصلية، نستخدم (8.50). وباستخدام البيانات في الجدول (٨ ـ٥)، نحصل على:

$$b_1=\left(\frac{s_Y}{s_1}\right)b_1'=\frac{62.049}{116.83}(0.9339)=0.496$$

$$b_2=\left(\frac{s_Y}{s_2}\right)b_2'=\frac{62.049}{730.64}(0.1083)=0.00920$$

$$b_0=\overline{Y}-b_1\overline{X}_1-b_2\overline{X}_2=150.60-.946(241.73)-.00920(2,961.9)=3.45$$

ولذلك تكون دالة الانحدار المقدَّرة لنموذج الانحدار المتعدد بمتغيراته الأصلية:

$$\hat{Y}=3.45+0.496X_1+0.00920X_2$$

وهي دالة الانحدار التوفيقية نفسها التي حصلنا عليها في الفصل السابع. ولا يمكن هنا مقارنة b_1 و b_2 بصورة مباشرة، لأن X_1 مقاسة بوحدات هي آلاف الأشخاص و X_2 مقاسة بوحدة الدولار.

وتُفسّر معاملات الانحدار المعيارية $b'_1 = 0.9339$ و $b'_2 = 0.1083$ أحيانا، كدليل على أن للمجتمع الهدف X_1 تأثير على المبيعات أكبر بكثير من الدخل الفردي المصرح به X_2 لأن b'_1 أكبر بكثير من b'_2 وعلى أي حال، وكما سنرى في الفقرة القادمة، يجب أن يكون المرء حذرا في تفسير معاملات الانحدار سواء أكانت معيارية أم لا. والسبب هو أنه عندما تكون المتغيرات المستقلة مرتبطة، فيما بينها، كما هو الحال هنا، فإن معاملات الانحدار تتأثر بالمتغيرات المستقلة الأخرى في النموذج. وفي بيانات شركة زارثان، نجد أن الارتباط بين X_1 و X_2 هو $r_{12} = 0.569$.

ووجود ارتباطات بين المتغيرات المستقلة ليس السبب الوحيد الذي يؤثر في مقدار معاملات الانحدار المعيارية وإنما للفجوات بين مشاهدات المتغيرات المستقلة أثرها هي الأخرى، وأحيانا تكون مثل هذه الفجوات كيفية تماما.

وبالتالي، فليس من الحكمة، عادة تفسير مقادير معاملات الانحدار المعيارية وكأنها تعكس الأهمية النسبية للمتغيرات المستقلة.

تعليقات

١ـ تقدم بعض حزم الحاسب كلا من معاملات الانحدار b_k للنموذج بمتغيراته الأصلية بالإضافة إلى المعاملات المعيارية b'_k وأحيانا، تُعطى هذه الأخيرة في المخرجات تحت عنوان معاملات بيتا.

٢ـ تبين بعض المخرجات الحاسوبية مقدار محدد مصفوفة الارتباط للمتغيرات X. وتتضمن القيمة القريبة من الصفر لهذا المحدد درجة عالية من الاقتران الخطّي بين المتغيرات X ، وزخما عاليا لأخطاء التدوير.

٣ـ عندما تُوسَّع مصفوفة الارتباط للمتغيرات X لتضم سطرا وعمودا خاصين بالمتغير Y، تُدعى مصفوفة الارتباط. وتبين مصفوفة الارتباط معاملات الارتباط لجميع الأزواج الممكنة من المتغير التابع والمتغيرات X. وهذه المعلومات مفيدة لأغراض متنوعة ـ على سبيل المثال، في اختيار المتغيرات المستقلة النهائية التي سيشملها النموذج.

ويعرض العديد من برامج الحاسب مصفوفة الارتباط في مطبوعة المخرجات. ومصفوفة الارتباط، في حالة متغيرين مستقلين، هي:

$$\begin{bmatrix} 1 & r_{Y1} & r_{Y2} \\ r_{Y1} & 1 & r_{12} \\ r_{Y2} & r_{12} & 1 \end{bmatrix}$$

وبما أن مصفوفة الارتباط متناظرة، فيحذف، في الغالب، القطاع المثلث الأدنى (أو الأعلى) من العناصر في مُخرجات الحاسب.

٤- من الممكن استخدام تحويل الارتباط في حزم الحاسب التي لاتسمح بانحدار عبر المبدأ، وذلك لأن b_0' ستكون دائما مساوية للصفر في بيانات حُولت بهذه الطريقة، وستكون معاملات الانحدار الأخرى صحيحة أيضا.

٥- سيقود استخدام المتغيرات المعيارية في (8.40a و8.40b)، دون اللجوء إلى تعديلات تحويل الارتباط في (8.41)، إلى معاملات الانحدار المعيارية نفسها، المعطاة في (8.49a)، لمتغيرات خضعت لتحويل الارتباط. إلا أن عناصر المصفوفة **X'X** سوف لاتكون عندئذ محدودة بين 1- و 1+.

(٨-٥) الخطّية المتعددة وتأثيراتها

في تحليل الانحدار المتعدّد، يهتم المرء غالبا بطبيعة وأهمية العلاقات بين المتغيرات المستقلة والمتغير التابع. ومن بين الأسئلة التي كثيرا ماتُثار:

١- ما هي الأهمية النسبية لتأثيرات المتغيرات المستقلة المختلفة؟

٢- ما هو مقدار تأثير متغير مستقل بعينه على المتغير التابع؟

٣- هل يمكن شطب أي من المتغيرات المستقلة من النموذج لأن تأثيره على المتغير التابع هو تأثير طفيف؟

٤- هل ينبغي النظر في إمكانية ضم أية متغيرات مستقلة، لم يشملها النموذج بعد، إلى النموذج؟

وإذا كانت المتغيرات المستقلة التي يشملها النموذج: (١) غير مرتبطة فيما بينها و(٢) غير مرتبطة مع أية متغيرات مستقلة أخرى تتصل بالمتغير التابع ولكنها ملغاة من

النموذج، فيمكن إعطاء أجوبة بسيطة نسبيا على هذه الأسئلة ومن سوء الطالع، تميل المتغيرات المستقلة في العديد من الدراسات غير التجريبية في الأعمال، الاقتصاد، والعلوم الاجتماعية والبيولوجية، إلى أن تكون مرتبطة فيما بينها ومرتبطة مع متغيرات أخرى ذات صلة بالمتغير التابع وغير مشمولة في النموذج. وعلى سبيل المثال، في انحدار نفقات الطعام لأسرة على المتغيرات المستقلة: دخل الأسرة، توفيرات الأسرة، وعمر رب الأسرة، ستكون المتغيرات المستقلة مرتبطة فيما بينها. وأكثر من ذلك، ستكون المتغيرات المستقلة مرتبطة أيضا بمتغيرات اجتماعية ـ اقتصادية غير مشمولة في النموذج ولها تأثيرها عل نفقات طعام الأسرة، مثل حجم الأسرة.

وعندما تكون المتغيرات المستقلة مرتبطة فيما بينها يُقال أنه يوجد ارتباط داخلي أو خطّية متعددة فيما بينها. (وأحيانا نحتفظ بالمصطلح الأخير لتلك الحالات التي يكون الارتباط فيها، بين المتغيرات المستقلة، عاليا جدا). وسنستطلع الآن مشكلات متنوعة على صلة متبادلة فيما بينها ويخلقها وجود الخطّية المتعددة بين المتغيرات المستقلة. وعلى أي حال، فسنتعرض أولا للحالة التي تكون المتغيرات المستقلة فيها غير مرتبطة.

جدول (٨ ـ٦) بيانات مثال إنتاجية طاقم عمل بمتغيرات مستقلة غير مرتبطة.

التجربة i	حجم الطاقم X_{i1}	العلاوات المدفوعة X_{i2}	درجة إنتاجية الطاقم Y_i
1	4	$2	42
2	4	2	39
3	4	3	48
4	4	3	51
5	6	2	49
6	6	2	53
7	6	3	61
8	6	3	60

التأثيرات عندما تكون المتغيرات المستقلة غير مرتبطة

يتضمن الجدول (٨ـ٦) بيانات تجربة على نطاق ضيق حول تأثير حجم طاقم عمل (X_1) ومستوى العلاوات (X_2) على درجة إنتاجية الطاقم (Y). ومن السهل تبيان أن X_1 و X_2 غير مرتبطين هنا، أي أن $r_{12}^2=0$ حيث يرمز r_{12}^2 لمعامل التحديد البسيط بين X_1 و X_2. ويتضمن الجدول (٨ـ٧) أ دالة الانحدار التوفيقية وجدول تحليل التباين عندما يتضمن النموذج كلا من X_1 و X_2 ويتضمن الجدول (٨-٧)ب المعلومات نفسها

عندما يكون فقط مشمولا في النموذج، ويتضمن الجدول (٨ ـ٧)جـ هذه المعلومات عندما يكون X_2 .بمفرده في النموذج.

والناحية المهمة الجديرة بالملاحظة في الجدول (٨ ـ٧) هي أن معامل الانحدار b_1=5.375 الخاص بـ X_1 يبقى نفسه سواء تضمن النموذج X_1 فقط أم تضمن كلا من المتغيرين المستقلين. والأمر نفسه صحيح من أجل b_2 = 9.250 وهذا نتيجة لكون المتغيرين المستقلين غير مرتبطين.

جدول (٨ ـ٧) جداول تحاين لمثال إنتاجية طاقم عمل بمتغيرات مستقلة غير مرتبطة

(أ) انحدار Y على X_1 و X_2

$$\hat{Y} = 0.375 + 5.375X_1 + 9.250X_2$$

مصدر التغير	SS	df	MS
الانحدار	402.250	2	201.125
الخطأ	17.625	5	3.525
المجموع	419.875	7	

(ب) انحدار Y على X_1

$$\hat{Y} = 23.500 + 5.375X_1$$

مصدر التغير	SS	df	MS
الانحدار	231.125	1	231.125
الخطأ	188.750	6	31.458
المجموع	419.875	7	

(جـ) انحدار Y على X_2

$$\hat{Y} = 27.250 + 9.250X_2$$

مصدر التغير	SS	df	MS
الانحدار	171.125	1	171.125
الخطأ	248.750	6	41.458
المجموع	419.875	7	

وهكذا، عندما تكون المتغيرات المستقلة غير مرتبطة، فإن التأثيرات المنسوبة لها، بواسطة نموذج انحدار من المرتبة الأولى، تبقى نفسها بصرف النظر عن أية متغيرات مستقلة يشملها النموذج. وهذه حجة قوية في صالح التجارب الخاضعة للتحكم وذلك حيثما تكون مثل هذه التجارب ممكنة، باعتبار أن التحكم التجريبي يسمح بجعل المتغيرات المستقلة غير مرتبطة.

والناحية المهمة الأخرى في الجدول (٨-٧) تتصل بمجاميع مربعات الخطأ، إذ نلاحظ من الجدول (٨-٧) أن مجموع المربعات الإضافي $SSR(X_1 | X_2)$ يساوي مجموع مربعات الانحدار $SSR(X_1)$ عندما يقتصر نموذج الانحدار على المتغير X_1 :

$$SSR(X_1 | X_2) = SSE(X_2) - SSE(X_1, X_2)$$
$$= 248.750 - 17.625 = 231.125$$
$$SSR(X_1) = 231.125$$

وبصورة مماثلة، مجموع المربعات الإضافي $SSR(X_2 | X_1)$ يساوي مجموع مربعات الانحدار $SSR(X_2)$ عندما يقتصر نموذج الانحدار على المتغير X_2 :

$$SSR(X_2 | X_1) = SSE(X_1) - SSE(X_1, X_2)$$
$$= 188.750 - 17.625 = 171.125$$
$$SSR(X_2) = 171.125$$

وبصورة عامة، عندما يكون متغيران مستقلان (أو أكثر) غير مرتبطين (مرتبطة) فإن المساهمة الهامشية لمتغير مستقل واحد في تخفيض مجموع مربعات الخطأ، مع وجود المتغيرات الأخرى في النموذج، تبقى نفسها بالضبط كما لو أن النموذج كان يقتصر على ذلك المتغير المستقل بمفرده.

ملاحظة

لتبيان أن معامل الانحدار لـ X_1 لايتغير عند إضافة X_2 إلى نموذج الانحدار في الحالة التي يكون فيها X_1 و X_2 غير مرتبطين، لنتأمل العبارة الجبرية لـ b_1 في نمودج الانحدار المتعدد بمتغيرين مستقلين:

$$b_1 = \frac{\dfrac{\Sigma(X_{i1} - \overline{X}_1)(Y_i - \overline{Y})}{\Sigma(X_{i1} - \overline{X}_1)^2} - \left[\dfrac{\Sigma(Y_i - \overline{Y})^2}{\Sigma(X_{i1} - \overline{X}_1)^2}\right]^{1/2} r_{Y2} r_{12}}{1 - r_{12}^2} \qquad (8.53)$$

حيث يرمز r_{Y2} ، كما سبق، لمعامل الارتباط البسيط بين Y و X_2 ، ويرمز r_{12} لمعامل الارتباط البسيط بين X_1 و X_2.

وإذا كان X_1 ، X_2 غير مرتبطين $r_{12} = 0$ ، فإن (8.53) تُختزل إلى:

$$b_1 = \frac{\sum(X_{i1} - \bar{X}_1)(Y_i - \bar{Y})}{\sum(X_{i1} - \bar{X}_1)^2} \qquad (8.53a)$$

ولكن (8.53a) هو مقدّر الميل لانحدار خطّي بسيط لِ Y على X_1 وفقا لِ (2.10a).

وبالتالي، عندما يكون X_1 و X_2 غير مرتبطين فإن إضافة X_2 إلى نموذج الانحدار لايغير من معامل انحدار X_1 ، وفي المقابل، فإن إضافة X_1 إلى نموذج الانحدار لايغير من معامل الانحدار لـ X_2.

طبيعة المشكلة عندما تكون المتغيرات المستقلة مرتبطة تماما

لرؤية الطبيعة الأساسية لمشكلة الخطّية المتعددة، سنستخدم مثالا بسيطا يكون فيه المتغيران المستقلان مرتبطين تماما، وتشير البيانات في الجدول (٨-٨) إلى عيّنة من أربع مشاهدات لمتغير تابع ولمتغيرين مستقلين. وقد طُلب من السيد (أ) توفيق دالة الانحدار المتعدد:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (8.54)$$

وقد عاد بعد وقت قصير بدالة الانحدار التوفيقية:

$$\hat{Y} = -87 + X_1 + 18X_2 \qquad (8.55)$$

وكان فخورا بأن دالة الاستجابة قد وافقت البيانات تماما. والقيم التوفيقية مبينة في الجدول (٨-٨).

وقد اتفق أن طُلب من السيد (ب) أن يقوم أيضا بتوفيق دالة الاستجابة (8.54) للبيانات نفسها وقد حصل بفخر على:

$$\hat{Y} = -7 + 9X_1 + 2X_2 \qquad (8.56)$$

وتتفق دالة استجابته أيضا اتفاقا تاما مع البيانات، كما هو مبين في الجدول (٨-٨).

وفي الحقيقة، يمكن تبيان أن مالانهاية لـه مـن دوال الاستجابة ستتفق تماما مع البيانات في الجدول (٨-٨) والسبب هو أن المتغيرين المستقلين X_1 و X_2 على صلة تامة ببعضهما وفقا للعلاقة:

$$X_2 = 5 + 0.5X_1 \tag{8.57}$$

جدول (٨-٨) مثال متغيرات مستقلة مرتبطة تماما

الحالة i	X_{i1}	X_{i2}	Y_i	القيم التوفيقية الدالة (8.55)	القيم التوفيقية الدالة (8.56)
1	2	6	23	23	23
2	8	9	83	83	83
3	6	8	63	63	63
4	10	10	103	103	103

دوال الاستجابة

$$\hat{Y} = -87 + X_1 + 18X_2 \quad (8.55)$$

$$\hat{Y} = -7 + 9X_1 + 2X_2 \quad (8.56)$$

لاحظ بعناية أن دالتي الاستجابة التوفيقيتين في (8.55) و(8.56) هما سطحا استجابة مختلفان كليا. ومعاملات الانحدار مختلفة، والقيم التوفيقية ستختلف عندما لاتتبع X_1 وX_2 العلاقة (8.57). وعلى سبيل المثال، القيمة التوفيقية لدالة الاستجابة (8.55) عندما يكون $X_1 = 5$ و $X_2 = 5$ هي:

$$\hat{Y} = -87 + 5 + 18(5) = 8$$

بينما القيمة التوفيقية لدالة الاستجابة (8.56) هي:

$$\hat{Y} = -7 + 9(5) + 2(5) = 48$$

وهكذا عندما يكون X_1 وX_2 على صلة تامة، كما في مثالنا، فالبيانات لا تتضمن أية مركّبة خطأ عشوائي، وسيؤدي العديد من دوال الاستجابة المختلفة إلى القيم نفسها المتفقة تماما مع المشاهدات، وإلى القيم نفسها المتفقة تماما مع أية تركيبات أخرى (X_1, X_2) تتبع العلاقة بين X_1 و X_2. ومع ذلك فإن دالتي الاستجابة هاتين ليستا متطابقتين وستؤديان إلى قيم توفيقية مختلفة لتركيبات (X_1, X_2) لاتتبع العلاقة بين X_1 وX_2.

وهناك مضمونان رئيسان لهذا المثال هما:

١ـ لاتكبح العلاقة التامة بين X_1 و X_2 مقدرتنا على الحصول على توفيق جيد للبيانات.

٢ـ بما أن العديد من دوال الاستجابة المختلفة تقدم التوفيق الجيد نفسه، فلا يمكن للمرء أن يفسر أية مجموعة من معاملات الانحدار كانعكاس لتأثيرات المتغيرات المستقلة المختلفة، وهكذا ففي دالة الاستجابة (8.55) لا يتضمن كون $b_1 = 1$ و$b_2 = 18$ أن X_2 هو المتغير المستقل الرئيس وأن X_1 يلعب دورا ثانويا، ذلك لأن دالة الاستجابة (8.56) تقدم توفيقا على المستوى نفسه من الجودة ولمعاملات انحدارها مقادير، هي بالمقارنة مع المعاملات السابقة، معاكسة تماما.

تأثيرات الخطّية المتعدّدة

في الواقع العملي، نادرا مانجد متغيرات مستقلة على علاقة تامة ببعضها أو بيانات لاتتضمن مركّبة خطأ عشوائي، ومع ذلك تبقى المضامين التي لاحظناها لتوّنا في مثالنا النموذجي وثيقة الصلة بموضوعنا.

١ـ بصورة عامة، لاتكبح حقيقة أن تكون بعض المتغيرات المستقلة أو جميعها مرتبطة فيما بينها، قدرتنا على الحصول على توفيق جيد. ولاتنزع إلى التأثير في استقراءات حول متوسط الاستجابة، أو تنبؤات بمشاهدات جديدة، شريطة أن تتم هذه الاستقراءات ضمن منطقة المشاهدات (يقدم الشكل (٧-٩) على الصفحة ٣٣٢ توضيحا لفكرة منطقة المشاهدات في حالة متغيرين مستقلين).

٢ـ العديد من دوال الانحدار المختلفة التي تقدم، في مثالنا النموذجي، توفيقات للبيانات على المستوى نفسه من الجودة، يقابلها في الحياة العملية أن معاملات الانحدار المقدّرة تميل إلى أن يكون لها تشتت معاينة كبير عندما تكون المتغيرات المستقلة على درجة عالية من الارتباط، وهكذا تميل معاملات الانحدار المقدّرة إلى أن تتغير تغيرا واسعا من عيّنة إلى عيّنة وذلك عندما تكون المتغيرات المستقلة عالية الارتباط. وكنتيجة

لذلك، لايمكن أن تتوافر لنا إلا معلومات غير دقيقة عن حقيقة معاملات الانحدار كـلّ بمفردها. وفي الحقيقة، يمكن أن تكون كل من معاملات الانحدار المقـدرة بمفردهـا غـير مهمة إحصائيا مع أن هناك بالتأكيد علاقة إحصائية قائمة بـين المتغـير التـابع ومجموعـة المتغيرات المستقلة.

٣- التفسير الشائع لمعاملات الانحـدار كقيـاس للتغـير في القيمـة المتوقعـة للمتغـير التابع عندما يزداد المتغير المستقل المقابل بوحدة واحدة، مع بقاء جميع المتغيرات المستقلة الأخرى ثابتة، يصبح تفسيرا غير قابل للتطبيق تماما عند تواجد الخطّية المتعـددة. فبينمـا قد يكون من الممكن نظريا تغيير متغير مستقل واحد وإبقاء المتغيرات الأخرى ثابتة، إلاّ أنه قد لايكون ذلك ممكنا عمليا في حالة متغيرات مستقلة عالية الارتباط. وعلى سـبيل المثال، في نموذج انحدار للتنبؤ بإنتاج محصول من معدل هطول المطر وساعات الشـمس المشرقة، ستجعل العلاقة بين المتغيرين المستقلين أمر تغيير أحدهما مع بقـاء الآخـر ثابتـا أمرا غير واقعي. وبالتالي لايمكن ضمان التفسير البسـيط لمعـاملات الانحـدار علـى أنهـا قياس للتأثيرات الهامشية، عند وجود متغيرات مستقلة عالية الارتباط.

وسـنوضح هـذه التأثـيرات للخطّيـة المتعـددة بـالعودة إلى مثـال شـحوم الجسـم، فالبيانات الأساسية معطاة في الجدول (٨-١)، ونتائج الانحـدار لنمـاذج توفيـق مختلفـة مقدمة في الجدول (٨-٢). والمتغيران المستقلان X_1 و X_2 مرتبطان ارتباطا عاليا كما يمكن أن نـرى مـن مصفوفـة الارتبـاط للمتغـيرات X في الجـدول (٨-١)ب. ومعـامل الارتباط البسيط $r_{12} = 0.92$ ، وعلى الوجه الآخر، فإن X_3 لاترتبط مثل هذا الارتبـاط العالي بأي من X_1 و X_2 بمفرده، فمعاملات الارتبـاط هـي $r_{13} = 0.46$ و $r_{23} = 0.08$. (إلاّ أن X_3 عالية الارتباط بـ X_1 و X_2 معا، فمعامل التحديد المتعدد عند انحدار X_3 على X_1 و X_2 هو 0.998).

تأثيرات معاملات الانحدار. نلاحظ من الجـدول (٨-٢) أن معـامل الانحـدار لـ X_1 ، سماكة الجلد في عضلة مؤخر العضد الثلاثية، يتغـير بصـورة ملحوظـة وفقـا للمتغـيرات الأخرى التي يشملها النموذج:

المتغيرات في النموذج	b_1	b_2
X_1	0.8572	-
X_2	_	0.8565
X_1 , X_2	0.2224	0.6594
X_1 , X_2 , X_3	4.334	-2.857

والقصة نفسها بالنسبة لمعامل الانحدار الخاص بـ X_2. وفي الحقيقة، يتغير معامل الانحدار b_2 حتى في إشارته عندما يضاف X_3 إلى النموذج المتضمن لِ X_1 و X_2.

والنتيجة المهمة التي يجب استخلاصها هي: عندما تكون المتغيرات المستقلة مرتبطة، فإن معامل الانحدار لأي متغير مستقل يعتمد على المتغيرات المستقلة الأخرى، أيها مشمول في النموذج وأيها بقي خارج النموذج، وهكذا فإن معامل الانحدار لايعكس أي تأثير أصيل للمتغير المستقل بالذات على المتغير التابع ولكنه يعكس فقط تأثيرا جزئيا أو هامشيا، علما أن متغيرات مستقلة مرتبطة أخرى، أيا كانت، مشمولة في النموذج.

ملاحظة

وحقيقة أنه يمكن لمتغيرات مستقلة مرتبطة فيما بينها أن تؤثر في معاملات الانحدار في نموذج انحدار، عند حذفها من ذلك النموذج، تجد إيضاحا لها في حيرة محلل من إشارة معامل انحدار في نموذج انحدار قام بتوفيقه، فقد وجد في انحدار مبيعات شركة في منطقة على حجم سكان المنطقة، والدخل الفردي، وبعض المتغيرات المستقلة الأخرى أن فترة الثقة لمعامل انحدار حجم السكان تشير إلى أن هذا المعامل سالب، وكان ينبغي على المحلل أن يأخذ في الاعتبار بعض المتغيرات المستقلة المحذوفة عند بحثه عن تفسير لهذه النتيجة. وقد لاحظ مستشار أن المحلل لم يضع تغلغل المنافسين الرئيسين في السوق ضمن النموذج. وبما أن المنافس كان أكثر نشاطا وفاعلية في المناطق ذات العدد الكبير من السكان فقد هبطت مبيعات الشركة في هذه المناطق، وكانت نتيجة حذف هذا المتغير المستقل من النموذج معاملا سالبا لمتغير حجم السكان.

تأثيرات على مجاميع المربعات الإضافية. وكما في معامل الانحدار، فإن المساهمة الهامشية لمتغير مستقل في تخفيض مجموع مربعات الخطأ تتغير معتمدة على المتغيرات الأخرى، أيها مشمول في النموذج، وأيها خارج النموذج، وذلك عندما تكون المتغيرات المستقلة مرتبطة. وعلى سبيل المثال، يقدم الجدول (٨-٢) مجاميع المربعات الإضافية المثالية للمتغير X_1:

$$SSR(X_1) = 352.27$$
$$SSR(X_1 \mid X_2) = 3.47$$

وسبب كون $SSR(X_1 \mid X_2)$ في مثل هذا الصغر بالمقارنة مع $SSR(X_1)$ هو أن X_1 و X_2 على درجة عالية من الارتباط. وهكذا عندما تكون X_2 في نموذج الانحدار فإن المساهمة الهامشية لِ X_1 في تخفيض مجموع مربعات الخطأ هي مساهمة صغيرة نسبيا إذ ينطوي X_2 على كثير من المعلومات نفسها التي ينطوي عليها X_1.

والقصة نفسها نجدها في الجدول (٨-٢) من أجل X_2. وهنا $SSR(X_2 \mid X_1)$ وهو أصغر بكثير من $SSR(X_2) = 381.97$. والنتيجة المهمة هي هذه: عندما تكون المتغيرات المستقلة مرتبطة فلا يمكن نسبة أي مجموع مربعات بمفرده إلى متغير مستقل واعتباره انعكاسا لتأثير هذا المتغير المستقل في تخفيض التغير الكلي في Y. ويجب النظر إلى التخفيض في التغير الكلي، المنسوب إلى متغير مستقل، في سياق المتغيرات المستقلة الأخرى المشمولة في النموذج، وذلك حيثما كانت المتغيرات المستقلة مرتبطة.

ملاحظة

تؤثر الخطّية المتعددة أيضا في معاملات التحديد الجزئية عبر تأثيراتها على مجاميع المربعات الإضافية. وعلى سبيل المثال، نلاحظ من الجدول (٨-٢)، في مثال شحوم الجسم، أن X_1 على درجة عالية من الارتباط بـ Y:

$$r_{Y1}^2 = \frac{SSR(X_1)}{SSTO} = \frac{352.27}{495.39} = 0.71$$

إلا أن معامل التحديد الجزئي بين Y و X_1 ، عندما كان X_2 من حينه في نموذج الانحدار، أصغر بكثير:

$$r^2_{Y1.2} = \frac{SSR(X_1 | X_2)}{SSE(X_2)} = \frac{3.47}{113.42} = 0.03$$

وسبب صغر معامل التحديد الجزئي هنا هو، كما رأينا، الارتبـاط العـالي بـين X_1 وX_2 وبالتالي لايقدم X_1 إلاّ معلومات إضافية ضئيلة نسبيا فوق تلك التي يزودنا بها X_2.

تأثيرات على $s\{b_k\}$. ونلاحظ من الجـدول (٨-٢) كـم نزيـد في عـدم دقـة معـاملي الانحدار المقدّرين b_1 و b_2 كلما أضفنا مزيدا من المتغيرات المستقلة إلى نموذج الانحدار.

المتغيرات في النموذج	$s\{b_1\}$	$s\{b_2\}$
X_1	0.1288	-
X_2	-	0.1100
X_1, X_2	0.3034	0.2912
X_1, X_2, X_3	3.016	2.582

ومرة ثانية نجد أن الدرجة العالية من الخطّية المتعددة بين المتغيرات المستقلة هي المسؤولة عن تضخم متغيرية معاملات الانحدار المقدّرة.

تأثيرات على القيم التوفيقية وقيم التنبؤ. لاحظ في الجدول (٨-٢) أن الخطّية المتعددة المرتفعة بين المتغيرات المستقلة لاتمنع من الانخفاض المطرد لمتوسط مربعات الخطأ ، وهو يقيس متغيرية حدود الخطأ، وذلك كلما دخلت متغيرات إضافية إلى نموذج الانحدار:

المتغيرات في النموذج	*MSE*
X_1	7.95
X_1, X_2	6.47
X_1, X_2, X_3	6.15

وفضلا عن ذلك، فإن دقة القيم التوفيقية ضمن مدى المشاهدات على المتغيرات المستقلة لاتتآكل بإضافة متغيرات مستقلة إلى نموذج الانحدار. لنأخذ مسألة تقدير متوسط شحوم الجسم عندما يكون المتغير المستقل الوحيد ضمن النموذج هو سماكة الجلد في عضلة مؤخر العضد الثلاثية (X_1) وذلك من أجل $X_{h1} = 25.0$ فالقيمة التوفيقية وانحرافها المعياري المقدّر هما (الحسابات غير معطاة):

$$\hat{Y}_h = 19.93 \qquad s\{\hat{Y}_h\} = 0.632$$

وعند إضافة متغير مستقل آخر عالي الارتباط إلى النموذج ، وهو محيط الفخذ X_2، فإن تقدير متوسط شحوم الجسم وانحرافه المعياري المقدّر هما كما يلي من أجل $X_{h1} = 25.0$ و $X_{h2} = 50.0$:

$$\hat{Y}_h = 19.36 \qquad s\{\hat{Y}_h\} = 0.624$$

وهكذا فإن دقة تقدير متوسط الاستجابة هي على الدرجة نفسها من الجودة كدقة التقدير السابق، بالرغم من إضافة متغير مستقل ثان عالي الارتباط بالمتغير الأول. وقد حصل هذا الاستقرار في دقة تقدير متوسط الاستجابة بالرغم من حقيقة أن الانحراف المعياري المقدر لِ b_1 أصبح أكبر بكثير عندما أضيف X_2 إلى النموذج (جدول ٨-٢) والسبب الأساسي للاستقرار هو أن التغاير بين b_1 و b_2 سالب، وهو يشكل مضادا، قوي التأثير للزيادة في $s^2\{b_1\}$ عند تحديد قيمة $s^2\{\hat{Y}_h\}$ كما هو معطى في (7.71).

وعندما يتضمن النموذج جميع المتغيرات المستقلة الثلاثة فإن تقدير متوسط شحوم الجسم وانحرافه المعياري المقدّر يصبحان كما يلي من أجل $X_{h1} = 25.0$ ، $X_{h2} = 50.0$ و $X_{h3} = 29.0$:

$$\hat{Y}_h = 19.19 \qquad s\{\hat{Y}_h\} = 0.621$$

وهكذا فإن إضافة متغير مستقل ثالث عالي الارتباط بالمتغيرين المستقلين الأولين معا، لم يؤثر تأثيرا يذكر بدقة تقدير متوسط الاستجابة.

تأثيرات على اختبارات متزامنة لِ β_k. وإساءة الاستخدام التي لايمكن اعتبارها قليلة الحدوث في تحليل نماذج الانحدار المتعدد هي دراسة الإحصاءة t^* في (7.46b):

$$t^* = \frac{b_k}{s\{b_k\}}$$

من أجل كل معامل انحدار، واحدا بعد الآخر، وذلك لتقدير ما إذا كان $\beta_k = 0$ من أجل $k = 1,...,p-1$. وحتى عند استخدام طرق الاستقراء المتزامن، وقليلا ما تُستخدم، فلاتزال توجد مشاكل عندما تكون المتغيرات المستقلة عالية الارتباط فيما بينها.

لنفرض أننا نرغب في اختبار ما إذا كان $\beta_1 = 0$ و $\beta_2 = 0$ في نموذج انحدار مثال شحوم الجسم بمتغيرين مستقلين، المبين في الجدول (٨-٢)جـ. وبضبط مستوى المعنوية العائلي عند 0.05 نحتاج وفقا لطريقة بونفرّوني إلى القيام بكل من الاختبارين t عند مستوى معنوية 0.025. وبالتالي نحتاج إلى $t(.9875\ ;\ 17) = 2.46$ وبما أن إحصاءتي الاختبار t^* في الجدول (٨-٢)جـ لاتتجاوزا، بالقيمة المطلقة، 2.46 فنستنتج من الاختبارين المنفصلين أن $\beta_1 = 0$ و $\beta_2 = 0$. ومع ذلك فإن الاختبار F للفرضية $H_0: \beta_1 = \beta_2 = 0$ يستنتج H_a بأن المعاملتين لاتساويان الصفر في آن واحد. ويمكن رؤية هذا من الجدول (٨-٢)جـ، حيث نجد: $F^* = MSR / MSE = 192.72 / 6.74 = 29.8$ وهو يتجاوز بكثير $F(.95; 2, 17) = 3.59$.

وسبب هذه النتيجة، التي تبدو كمفارقة غير مقبولة، هو أن الاختبار t^* هو اختبار هامشي، كما رأينا في (8.15) من منظور أسلوب الاختبار الخطّي العام. وهكذا، عندما يكون X_1 و X_2 مرتبطين ارتباطا عاليا فإن $SSR(X_1 \mid X_2)$ صغير، مما يشير إلى أن X_1 لا يقدم الكثير من المعلومات الإضافية زيادة عما يقدمه X_2 الموجود من حينه في النموذج؛ وهذا يقودنا بالتالي إلى النتيجة $\beta_1 = 0$. وبصورة مماثلة نُقاد إلى النتيجة $\beta_2 = 0$ هنا لأن $SSR(X_2 | X_1)$ صغير، مما يشير إلى أن X_2 لاينطوي على كثير من المعلومات الإضافية عندما يكون X_1 من حينه في النموذج. ولكن اختباري التأثيرين الهامشيين لـ X_1 و X_2 معا ليسا مكافئين لاختبار ما إذا كانت هناك علاقة انحدار بين Y والمتغيرين المستقلين X_1 و X_2. والسبب هو أن النموذج المخفض لكل من الاختبارين المنفصلين يتضمن المتغير المستقل الآخر، بينما لايتضمن النموذج المخفض لاختبار ما إذا كان كل من $\beta_1 = 0$ و $\beta_2 = 0$ أيا من المتغيرين المستقلين. ويبين الاختبار المناسب، وهو الاختبار F وجود علاقة انحدار لاريب فيها بين Y، X_1 و X_2.

وسنواجه المفارقة نفسها في الجدول (٨-٢)د لنموذج انحدار بثلاثة متغيرات مستقلة، إذا قمنا بثلاثة اختبارات متزامنة لمعاملات الانحدار عند مستوى معنوية عائلي 0.05.

ملاحظة

رأينا آنفا أنه يمكن لمجموعة من المتغيرات المستقلة أن تكون على صلة بالمتغير التابع، ومع ذلك فإن جميع الاختبارات المنفردة لمعاملات الانحدار ستؤدي إلى نتيجة أنها مساوية للصفر، وذلك بسبب الخطّية المتعددة بين المتغيرات المستقلة. وهذه النتيجة المحيّرة في الظاهر هي أيضا نتيجة ممكنة تحت ظروف خاصة وحيث لاتوجد خطية متعددة بين المتغيرات المستقلة. وعلى أي حال، فإنه ليس من المنتظر وجود مثل تلك الظروف في التطبيق العملي.

تشخيصات الخطّية المتعددة وتدابير علاجية

وكما رأينا، يمكن أن يكون للخطّية المتعددة بين المتغيرات المستقلة عواقب وخيمة فيما يتعلق بتفسير واستخدام نموذج انحدار قمنا بتوفيقه. والأداة التشخيصية التي نستعرضها هنا للتعرف على الخطّية المتعددة ـ ونعني، معاملات الارتباط البسيط بين كل زوج من أزواج المتغيرات المستقلة ـ غالبا ماتكون مفيدة. وعلى أي حال، فهناك ظروف تتواجد فيها خطّية متعددة خطرة دون أن تفصح عنها معاملات الارتباط بين أزواج المتغيرات. ونقدم في الفصل الحادي عشر أداة أكثر مقدرة على كشف وجود خطّية متعددة خطرة. وسنناقش هناك أيضا عددا من التدابير العلاجية لتقليل تأثيرات الخطّية المتعددة.

تعليقات

١ـ لوحظ في الفقرة (٨-٤) أن المحدد القريب من الصفر لـ $\mathbf{X'X}$ هو مصدر مهم من مصادر الأخطاء الجدية للتدوير في نتائج المربعات الدنيا. وإحدى نتائج الخطّية المتعددة الشديدة هي أن تجعل هذا المحدد قريبا من الصفر، وهكذا يمكن لمعاملات الانحدار أن تخضع، تحت الخطّية المتعددة الشديدة، لأخطاء تدوير كبيرة بالإضافة إلى تباينات معاينة كبيرة. وبالتالي فإنه من المستحسن، على وجه الخصوص، استخدام تحويل الارتباط (8.41) عند توفيق نموذج انحدار، وذلك في حال وجود الخطّية المتعددة.

٢- وكما أن الارتباط العالي بين المتغيرات المستقلة يميل إلى جعل معاملات الانحدار المقدّرة غير دقيقة (غريبة الأطوار من عيّنة إلى عيّنة) فكذلك تميل معاملات الارتباط الجزئي بين المتغير التابع وكل من المتغيرات المستقلة إلى أن تصبح غريبة الأطوار من عيّنة إلى عيّنة عندما تكون المتغيرات المستقلة مرتبطة ارتباطا عاليا.

٣- يمكن بسهولة رؤية أثر الارتباط فيما بين المتغيرات المستقلة على الانحرافات المعيارية لمعاملات الانحدار المقدّرة، وذلك عند تحويل المتغيرات في النموذج مستخدمين تحويل الارتباط (8.41). فلنعتبر النموذج من المرتبة الأولى بمتغيرين مستقلين:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \tag{8.58}$$

ويصبح هذا النموذج في المتغيرات بعد التحويل (8.41):

$$Y_i' = \beta_1' X_{i1}' + \beta_2' X_{i2}' + \varepsilon_i' \tag{8.59}$$

والمصفوفة $(\mathbf{X'X})^{-1}$ لهذا النموذج المعياري معطاة في (8.51c):

$$(\mathbf{X'X})^{-1} = \mathbf{r}_{XX}^{-1} = \frac{1}{1-r_{12}^2}\begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \tag{8.60}$$

وبالتالي فإن مصفوفة التغاير لمعاملات الانحدار المقدّرة، باستخدام (7.41) هي:

$$\sigma^2\{\mathbf{b}\} = (\sigma')^2 \mathbf{r}_{XX}^{-1} = (\sigma')^2 \frac{1}{1-r_{12}^2}\begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \tag{8.61}$$

حيث $(\sigma')^2$ هو تباين حد الخطأ للنموذج المعياري (8.59).

وهكذا نجد التباين نفسه هنا لمعاملي الانحدار المقدرين b_1' و b_2':

$$\sigma^2\{b_1'\} = \sigma^2\{b_2'\} = \frac{(\sigma')^2}{1-r_{12}^2} \tag{8.62}$$

ويصبح هذا التباين أكبر كلما ازداد الارتباط بين X_1 و X_2. وفي الحقيقة، عندما يقترب X_1 وX_2 من ارتباط تام (أي عندما يتقارب r_{12}^2 إلى الواحد)، يصبح تباينا b_1' و b_2' كبيرين بلا حدود.

(٨ ـ ٦) صياغة مصفوفية لاختبار خطّي عام

الإجراءات الملخّصة في الفقرة (٨-٢) حول اختبارات تتعلق بمعاملات الانحدار هي إجراءات مناسبة. إذ يمكن استخدام مجاميع مربعات إضافية حيثما نرغب في اختبار أن بعض معاملات الانحدار مساوية للصفر، وفيما عدا ذلك يمكن توفيق النموذجين التام والمخفض عند إجراء الاختبار الخطّي العام حول معاملات الانحدار.

ومن وقت لآخر، على أي حال، يكون من الضروري تنفيذ الاختبار الخطي العام في صيغة مصفوفية، مثل أنواع معينة من الاختبارات في تحليل التباين. وسنشرح الآن كيف يمكن تمثيل إحصاءة الاختبار الخطي العام (3.69):

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

في صيغة مصفوفة.

نموذج تام

نموذج الانحدار التام المتضمن لـ $p - 1$ من متغيرات التنبؤ مُعطي في (7.18):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (8.63)$$

وسنرمز الآن لمقدّرات المربعات الدنيا للنموذج التام بـ $\mathbf{b}_F$ وهي معطاة، كما سبق، بالعلاقة (7.21):

$$\mathbf{b}_F = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \qquad (8.64)$$

وأيضا، مجموع مربعات الخطأ معطى بالعلاقة (7.30):

$$SSE(F) = (\mathbf{Y} - \mathbf{X}\mathbf{b}_F)'(\mathbf{Y} - \mathbf{X}\mathbf{b}_F) = \mathbf{Y}'\mathbf{Y} - \mathbf{b}_F'\mathbf{X}'\mathbf{Y} \qquad (8.65)$$

ويترافق معه $df_F = n - p$ درجة من الحرية.

عبارة الفرضية H_0

تُمثل فرضية اختبار خطّي H_0 بالصيغة المصفوفية كما يلي:

$$H_0: \underset{s\times p}{\mathbf{C}}\ \underset{p\times 1}{\mathbf{B}} = \underset{s\times 1}{\mathbf{h}} \qquad (8.66)$$

حيث $\mathbf{C}$ مصفوفة معينة $s \times p$ ورتبتها s، و $\mathbf{h}$ متجه معين $s \times 1$.

مثال (١). يتضمن نموذج الانحدار متغيري X ، ونرغب في اختبار $H_0: \beta_1 = 0$

وعندئذ:

$$\underset{1\times3}{\mathbf{C}} = [0 \quad 1 \quad 0] \quad \underset{1\times1}{\mathbf{h}} = [0]$$

ولدينا

$$H_0:\ \mathbf{C}\beta = \begin{bmatrix}0 & 1 & 0\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix} = [0]$$

أو

$$H_0\colon \beta_1 = 0$$

مثال (٢). يتضمن نموذج الانحدار متغيرين مستقلين X ، ونرغب في اختبار $H_0\colon \beta_1 = \beta_2 = 0$، فعندئذ:

$$\underset{2\times3}{\mathbf{C}} = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} \qquad \underset{2\times1}{\mathbf{h}} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

ولدينا:

$$H_0\colon \mathbf{C}\beta = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

أو $H_0\colon \beta_1 = 0, \quad \beta_2 = 0$

مثال (٣). يتضمن نموذج الانحدار ثلاثة متغيرات مستقلة X، ونرغب في اختبار الفرضية $H_0\colon \beta_1 = \beta_2$، فعندئذ:

$$\underset{1\times4}{\mathbf{C}} = [0 \quad 1 \quad -1 \quad 0] \quad \underset{1\times1}{\mathbf{h}} = [0]$$

ولدينا:

$$H_0\colon \mathbf{C}\beta = \begin{bmatrix}0 & 1 & -1 & 0\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{bmatrix} = [0]$$

أو $H_0\colon \beta_1 - \beta_2 = 0$

نموذج مخفض

النموذج المخفض هو:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \qquad \text{حيث :} \quad \mathbf{C}\beta = \mathbf{h} \qquad (8.67)$$

ويمكن تبيان أن مقدّرات المربعات الدنيا تحت النموذج المخفض، وسنرمز لها بـ b_R هي:

$$\mathbf{b}_R = \mathbf{b}_F - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\mathbf{b}_F - \mathbf{h}) \qquad (8.68)$$

ومجموع مربعات الخطأ هو:

$$SSE(R) = (\mathbf{Y} - \mathbf{X}\mathbf{b}_R)'(\mathbf{Y} - \mathbf{X}\mathbf{b}_R) \qquad (8.69)$$

ويترافق معه $df_R = n - (p - s)$ درجة من الحرية.

إحصاءة اختبار

يمكن تبيان أنه يمكن التعبير عن الفرق $SSE(R) - SSE(F)$ كما يلي:

$$SSE(R) - SSE(F) = (\mathbf{C}\mathbf{b}_F - \mathbf{h})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\mathbf{b}_F - \mathbf{h}) \qquad (8.70)$$

ويترافق معه $df_R - df_F = (n - p + s) - (n - p) = s$ درجة من الحرية.

وبالتالي إحصاءة الاختبار هي:

$$F^* = \frac{SSE(R) - SSE(F)}{s} \div \frac{SSE(F)}{n-p} \qquad (8.71)$$

حيث $SSE(R) - SSE(F)$ معطى بالعلاقة (8.70) و$SSE(F)$ معطى بالعلاقة (8.65).

وللتثبت من أن درجات الحرية المترافقة مع $SSE(R) - SSE(F)$ هي s، لنتأمل الأمثلة الثلاثة السابقة:

١- في المثال (١) ، $s = 1$ ويتفق هذا مع عدد درجات الحرية في البسط في إحصاءة الاختبار (8.22).

٢- في المثال (٢)، $s = 2$ ويتفق هذا مع عدد درجات الحرية في البسط في إحصاءة الاختبار (8.20).

٣- في المثال (٣)، $s = 1$ ويتفق هذا مع عدد درجات الحرية في البسط في إحصاءة الاختبار للمثال المذكور على الصفحة ٣٥٩.

ملاحظة

يمكن استنباط مقدّرات المربعات الدنيا $\mathbf{b}_R$ تحت النموذج المخفض والمعطاة في (8.68)، بجعل معيار المربعات الدنيا $Q = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$ أصغر ما يمكن خاضعا للقيد $\mathbf{C}\beta - \mathbf{h} = 0$ مستخدمين مضاريب لاغرانج.

مسائل

(٨ ـ١) اكتب عدد درجات الحرية المترافقة مع كل من مجاميع المربعات الإضافية التالية:

(١) $SSR(X_1 | X_2)$ ؛

(٢) $SSR(X_2 | X_1, X_3)$ ؛

(٣) $SSR(X_1, X_2 | X_3, X_4)$ ؛

(٤) $SSR(X_1, X_2, X_3 | X_4, X_5)$ ؛

(٨ـ٢) بأي معنى يكون مجموع مربعات الانحدار $SSR(X_1)$ مجموع مربعات إضافي. اشرح.

(٨ـ٣) بالإشارة إلى مسألة **الصنف المفضل** (٧ـ٨).

أ ـ اكتب جدول تحليل التباين الذي يفكك مجموع مربعات الانحدار إلى مجاميع مربعات إضافية تترافق مع X_1 ومع X_2 ، علما أن X_1 معطاة.

ب ـ اختبر ما إذا كان يمكن شطب X_2 من نموذج الانحدار علما أن النموذج يتضمن X_1. استخدم إحصاءة الاختبار F^* ومستوى المعنوية 0.01. اعرض البديلين، قاعدة القرار، والنتيجة. ماهي القيمة $P-$ للاختبار؟

(٨ـ٤) بالإشارة إلى مسألة **شحن كيماويات** (٧ـ١٢).

أ ـ اكتب جدول التحليل التباين الذي يفكك مجموع مربعات الانحدار إلى مجاميع مربعات إضافية تترافق مع X_2 ومع X_1 علما أن X_2 مُعطى.

ب ـ اختبر ما إذا كان يمكن شطب X_1 من نموذج الانحدار علما أن النموذج يتضمن X_2. استخدم إحصاءة الاختبار F^* و $\alpha = 0.05$ اعرض البديلين، قاعدة القرار، والنتيجة. ماهي القيمة $P-$ للاختبار؟

جـ - هل $SSR(X_1) + SSR(X_2 | X_1)$ يساوي $SSR(X_2) + SSR(X_1 | X_2)$ هنا؟
هل يجب أن تكون الحال دائما كذلك؟

(٨-٥) بالإشارة إلى مسألة **ارتياح المريض** (٧-١٧).

أ - اكتب جدول تحليل التباين الذي يفكك مجموع مربعات الانحدار إلى مجاميع مربعات إضافية تترافق مع X_2 ؛ ومع X_1 ، علما أن X_2 مُعطى، ومع X_3 علما أن X_2 وX_1 معطيان.

ب - اختبر ما إذا كان يمكن شطب X_3 من نموذج الانحدار علما أن النموذج يتضمن X_1 وX_2 . استخدم إحصاءة الاختبار F^* ومستوى معنوية 0.025 اعرض البديلين وقاعدة القرار والنتيجة. ماهي القيمة -P للاختبار؟

(٨-٦) بالإشارة إلى مسألة **ارتياح المريض** (٧-١٧). اختبر ما إذا كان يمكن شطب X_2 وX_3 من نموذج الانحدار علما أن النموذج يتضمن X_1. استخدم $\alpha = 0.025$. اعرض البديلين، وقاعدة القرار، والنتيجة. ما هي القيمة - P للاختبار؟

(٨-٧) بالإشارة إلى مسألة **رواتب المتخصصين في الرياضيات** (٧-٢٠).

أ - اكتب جدول تحليل التباين الذي يفكك مجموع مربعات الانحدار إلى مجاميع مربعات إضافية تترافق مع X_2 ؛ ومع X_3 علما أن X_2 مُعطى؛ ومع X_1 ، علما أن X_2 و X_3 معطيان.

ب - اختبر ما إذا كان يمكن شطب X_1 من نموذج الانحدار علما أن النموذج يتضمن X_2 و X_3. استخدم إحصاءة الاختبار F^* ومستوى معنوية 0.01. اعرض البديلين، قاعدة القرار، والنتيجة. ما هي القيمة -P للاختبار؟

(٨-٨) بالإشارة إلى مسألتي رواتب **المتخصصين في الرياضيات** (٧-٢٠) و(٨-٧). اختبر ما إذا كان يمكن شطب X_1 و X_3 من نموذج الانحدار علما أن النموذج يتضمن X_2؛ استخدم $\alpha = 0.01$. اعرض البديلين، قاعدة القرار، والنتيجة، ماهي القيمة -P للاختبار ؟

(٨-٩) بالإشارة إلى مسألة **ارتياح المريض** (٧-١٧). اختبر ما إذا كان β_1=-1.0 و$\beta_2 = 0$ استخدم $\alpha = 0.025$ اعرض البديلين، والنموذجين التام والمخفض، وقاعدة القرار، والنتيجة.

(٨-١٠) بالإشارة إلى مسألة **رواتب المتخصصين** في الرياضيات (٧-٢٠) اختبر ما إذا كان $\beta_1 = \beta_3$؛ استخدم $\alpha = 0.01$ اعرض البديلين والنموذجين التام والمخفض وقاعدة القرار، والنتيجة.

(٨-١١) بالإشارة إلى مثال إنتاجية طاقم العمل على الصفحة ٣٧٧.

أ - احسب r^2_{Y1}، r^2_{Y2}، r^2_{12}، $r^2_{Y1.2}$، $r^2_{Y2.1}$ وR^2، اشرح مايقيسه كل معامل وفسّر نتائجك.

ب - هل تجد أن أيا من النتائج في (أ) خاص بسبب كون المتغيرين المستقلين غير مرتبطين؟

(٨-١٢) بالإشارة إلى مسألة **تفضيل صنف** (٧-٨). احسب r^2_{Y1}، r^2_{Y2}، r^2_{12}، $r^2_{Y1.2}$، $r^2_{Y2.1}$ وR^2، اشرح مايقيسه كل معامل وفسّر نتائجك.

(٨-١٣) بالإشارة إلى مسألة **شحنة الكيماويات** (٧-١٢). احسب r^2_{Y1}، r^2_{Y2}، r^2_{12}، $r^2_{Y1.2}$، $r^2_{Y2.1}$ وR^2، اشرح ما يقيسه كل معامل وفسّر نتائجك.

(٨-١٤) بالإشارة إلى مسألة **ارتياح المريض** (٧-١٧).

أ - احسب r^2_{Y1} ، $r^2_{Y1.2}$ و$r^2_{Y1.23}$. كيف تأثرت درجة الارتباط الخطّي بين Y وX_1 عندما أُضيفت متغيرات مستقلة أخرى إلى النموذج ؟

ب - قم بتحليل مماثل، كما في الجزء (أ) لدرجة الارتباط الخطّي بين Y و X_2 هل تماثل نتائجك هنا تلك التي حصلت عليها في الجزء (أ) من أجل Y و X_1 ؟

(٨-١٥) بالإشارة إلى مسألة **رواتب متخصصي الرياضيات** (٧-٢٠). احسب: r^2_{Y1}، r^2_{Y3}، r^2_{23}، $r^2_{Y1.2}$، $r^2_{Y1.23}$ وR^2 واشرح ما يقيسه كل معامل وفسّر نتائجك. كيف تتأثر درجة الارتباط الخطّي بين Y و X_1 عند إضافة متغيرات مستقلة أخرى إلى النموذج؟

(٨-١٦) بالإشارة إلى مسألة **تفضيل الصنف** (٧-٨).

أ - حوّل المتغيرات مستخدما تحويل الارتباط (8.41) وقم بتوفيق نموذج الانحدار المعياري (8.42).

ب - فسّر معامل الانحدار المعياري b_1'.

جـ - حوّل معاملات الانحدار المعيارية المقدَّرة مستخدما (8.50) لتعود بها إلى المعاملات الخاصة بتوفيق نموذج الانحدار في متغيراته الأصلية. تحقق أنها مساوية لتلك التي حصلت عليها في المسألة (٧-٨)أ.

(٨-١٧) بالإشارة إلى مسألة **شحنة كيماويات** (٧-١٢).

أ - حول المتغيرات مستخدما تحويل الارتباط (8.41) وقم بتوفيق نموذج الانحدار المعياري (8.42).

ب - احسب معامل التحديد بين المتغيرين المستقلين. هل من معنى هنا لاعتبار أن معاملات الانحدار المعيارية تعكس تأثير أحد المتغيرين المستقلين عندما نبقي الآخر ثابتا؟

جـ - حول معاملات الانحدار المعيارية المقدَّرة مستخدما (8.50) لتعود بها إلى المعاملات الخاصة بتوفيق نموذج الانحدار في متغيراته الأصلية. تحقق أنها مساوية لتلك التي حصلت عليها في المسألة (٧-١٢)جـ.

(٨-١٨) بالإشارة إلى مسألة **ارتياح المريض** (٧-١٧).

أ - حول المتغيرات مستخدما تحويل الارتباط (8.41) وقم بتوفيق نموذج الانحدار المعياري (8.42).

ب - احسب معاملات التحديد بين جميع الأزواج من المتغيرات المستقلة. هل تشير هذه إلى أن اعتبار معاملات الانحدار المعيارية هنا كمؤشر لتأثير أحد المتغيرات المستقلة عند إبقاء المتغيرات الأخرى مثبتة هو اعتبار لايخلو من معنى؟

جـ ـ حوِّل معاملات الانحدار المعيارية المقدّرة مستخدما (8.50) لتعود بها إلى المعاملات الخاصة بتوفيق نموذج الانحدار في متغيراته الأصلية، تحقق أنها مساوية لتلك التي حصلت عليها في المسألة (٧–١٧)ب.

(٨ـ١٩) بالإشارة إلى مسألة **رواتب متخصصي الرياضيات** (٧–٢٠).

أ ـ حول المتغيرات مستخدما تحويل الارتباط (8.41) وقم بتوفيق نموذج الانحدار المعياري (8.42).

ب ـ فسّر معامل الانحدار المعياري b_2' .

جـ ـ حول معاملات الانحدار المعيارية المقدرة باستخدام (8.50) لتعود بها إلى المعاملات الخاصة بتوفيق نموذج الانحدار في متغيراته الأصلية. تحقق أنها مساوية لتلك التي حصلت عليها في المسألة (٧–٢٠)ب.

(٨ـ٢٠) عرض متحدث في ندوة حول تحليل الانحدار التطبيقي: "في بيانات المسوح من ميادين الأعمال أو العلوم الاجتماعية، لايمكن، على وجه الخصوص، تجنب درجة ما من الخطية المتعددة "هل تنطبق هذه العبارة بالقدر نفسه على بيانات تجريبية؟

(٨ـ٢١) بالإشارة إلى مثال شركة زارثان على الصفحة ٣١٦ اقترح مدير المبيعات في الشركة أنه يمكن تحسين مقدرة النموذج على التنبوء تحسينا كبيرا ، إذا اضيفت النفقات التشجيعية إلى النموذج، باعتبار أنه من المعروف أن لهذه النفقات أثرا مهما على المبيعات. وقد خصصت الشركة الميزانية التشجيعية الإجمالية إلى المناطق بشكل يتناسب مع عدد سكانها. وهكذا فإن منطقة تتضمن 4.7 بالمائة من عدد السكان الإجمالي لمناطق الدراسة تتلقى 4.7 بالمائة من الميزانية التشجيعية الإجمالية، قوّم إقتراح مدير المبيعات.

(٨ـ٢٢) بالإشارة إلى مثال المتغيرات المستقلة المرتبطة تماما في الجدول (٨–٨):

أ ـ طوّر دالة استجابة أخرى، مشابهة لدالتي الاستجابة (8.55) و(8.56) تتفق مع البيانات اتفاقا تاما؟

ب ـ ما هو تقاطع العـدد اللانهـائي مـن سـطوح الاسـتجابة الـتي تتفـق مـع البيانات اتفاقا تاما؟.

(٨ـ٢٣) في تقرير لمحلل باحث، رفعه إلى الأستاذ المشرف عن تقدمه في البحث، جاء مايلي: "في نموذجنا بثلاثة متغيرات مستقلة المعد للتنبؤ بالمبيعات كانت جميع معاملات الانحدار المقدّرة مهمة إحصائيا، ونموذجنا الأوّلي الجديد بسبع متغيرات مستقلة، وهو يتضمن المتغيرات الثلاثة لنموذجنا المصغّر، هو نموذج أقل نجاحا لأن اثنين فقط من معاملات الانحدار السبعة مهمان إحصائيا، ومع ذلك ففي بعض الاستخدامات الأوّلية كان النموذج الموسّع يعطي تنبؤات بالمبيعات أكثر دقة من النموذج المصغّر. وأسباب هذه المفارقة قيد الدراسة الآن". علّق.

(٨ـ٢٤) كتب باحثان مايلي: "استخدم بحثنا نموذج انحدار متعدد. وقد تبين أن اثنين من المتغيرات المستقلة، وهما متغيران مهمان في نظريتنا، على درجة عالية من الارتباط في مجموعة البيانات المستخدمة. إلا أننا احتفظنا بالمتغيرين كليهما في النموذج لأن ارتفاع قيمة معامل التحديد المتعدد يجعل هذه الصعوبة غير ذات أهمية". علّق.

(٨ـ٢٥) بالإشارة إلى مسألة **تفضيل صنف** (٧-٨).

أ ـ قم بتوفيق نموذج الانحدار الخطّي البسيط من المرتبة الأولى (3.1) لإيجاد علاقة بين الرغبة في الصنف Y ومحتوى الرطوبة X_1، اكتب دالة الانحدار التوفيقية.

ب ـ قارن معامل الانحدار المقدّر لمحتوى الرطوبة الذي حصلت عليه في (أ) بالمعامل المقابل الذي حصلت عليه في المسألة (٧-٨)أ. ماذا وجدت ؟

جـ ـ هل $SSR(X_1)$ يساوي $SSR(X_1 | X_2)$ هنا؟ وإذا لم يكن الأمر كذلك، فهل الفرق مهم؟

د ـ أوجد مصفوفة الارتباط للمتغيرات X. ماهي الأعباء التي تلقيها هذه المصفوفة على نتائجك في الجزئين (ب) و(جـ)؟

(٨-٢٦) بالإشارة إلى مسألة **شحنة كيماويات** (٧-١٢).

أ - قم بتوفيق نموذج الانحدار الخطّي البسيط من المرتبة الأولى (3.1) لإيجاد علاقة بين عدد الدقائق اللازمة لاستلام شحنة Y وبين الوزن الاجمالي للشحنة X_2 اكتب دالة الانحدار التوفيقية.

ب - قارن معامل الانحدار المقدر للوزن الإجمالي للشحنة الذي وجدته في الجزء (أ) بالمعامل المقابل الذي حصلت عليه في المسألة (٧-١٢)جـ. ماذا تجد؟

جـ - هل $SSR(X_2)$ يساوي $SSR(X_2 \mid X_1)$ هنا ؟ وإذا لم يكن الأمر كذلك فهل الفرق مهم؟

د - أوجد مصفوفة الارتباط للمتغيرات X. ماهي الأعباء التي تلقيها هذه المصفوفة على نتائجك في الجزئين (ب) و(جـ)؟

(٨-٢٧) بالإشارة إلى مسألة **ارتياح المريض** (٧-١٧).

أ - قم بتوفيق نموذج الانحدار الخطّي من المرتبة الأولى (7.1) لإيجاد علاقة بين ارتياح المريض Y وعمر المريض X_1 وشدة المرض X_2. اكتب دالة الانحدار التوفيقية.

ب - قارن معاملي الانحدار المقدَّرين لعمر المريض وشدة المرض اللذين حصلت عليهما في الجزء (أ). بالمعاملين المقابلين اللذين حصلت عليهما في المسألة (٧-١٧)ب ماذا تجد؟

جـ - هل $SSR(X_1)$ يساوي هنا $SSR(X_1|X_3)$؟ هل $SSR(X_2)$ يساوي $SSR(X_2 \mid X_3)$؟

د - أوجد مصفوفة الارتباط للمتغيرات X. ماهي الأعباء التي تلقيها هذه المصفوفة على نتا ئجك في الجزئين (ب) و(جـ)؟

(٨-٢٨) بالاشارة إلى مسألة **رواتب المتخصصين في الرياضيات** (٧-٢٠).

أ - قم بتوفيق نموذج الانحدار الخطّي من المرتبة الأولى (7.1) لإيجاد علاقة بين الرواتب السنوية Y وبين نوعية النشر X_1 والخبرة X_2 اكتب دالة الانحدار المقدّرة.

ب - قارن معاملي الانحدار المقدَّرين لنوعية النشر والخبرة بالمعاملات المقابلة التي حصلت عليها في المسألة (٧-٢٠)ب ماذا تجد؟

جـ - هل $SSR(X_2)$ يساوي $SSR(X_2 | X_3)$ هنا؟ هل $SSR(X_1)$ و$SSR(X_1 \mid X_3)$ متساويان؟

د - أوجد مصفوفة الارتباط للمتغيرات X. ماهي الأعباء التي تلقيها هذه المصفوفة على نتائجك في الجزئين (ب) و(جـ)؟.

تمارين

(٨-٢٩) أ - عرِّف كلا من مجاميع المربعات الإضافية التالية:

(١) $SSR(X_5 | X_1)$ ؛ (٢) $SSR(X_3 , X_4 | X_1)$ ؛ (٣) $SSR(X_4 | X_1 ,X_2 , X_3)$ ؛

ب - في نموذج انحدار متعدد بخمس متغيرات مستقلة. ماهو مجموع المربعات الإضافي المناسب لاختبار ما إذا كان $\beta_5 = 0$ أم لا؟ ما إذا كان $\beta_2 = \beta_4 = 0$ أم لا؟.

(٨-٣٠) بيّن أن:

أ : $SSR(X_1,X_2,X_3,X_4) = SSR(X_1) + SSR(X_2,X_3 \mid X_1) + SSR(X_4 \mid X_1,X_2,X_3)$

ب: $SSR(X_1, X_2, X_3, X_4)=SSR(X_2, X_3) + SSR(X_1 | X_2, X_3) + SSR(X_4 \mid X_1, X_2, X_3)$

(٨-٣١) بالإشارة إلى مسألة **تفضيل صنف** (٧-٨).

أ - احدر Y على X_2 مستخدما نموذج الانحدار الخطّي البسيط (3.1) وأوجد الرواسب.

ب - احدر X_1 على X_2 مستخدما نموذج الانحدار الخطّي البسيط (3.1) وأوجد الرواسب.

جـ - احسب معامل الارتباط البسيط بين المجموعتين من الرواسب وبيّن أنها تساوي $r_{Y1.2}$.

(٨-٣٢) درس طالب جامعي يعمل في مخزن للثياب في المدينة الجامعية، يخدم زبائنه من الطلاب، العلاقة بين العلاوة الشهرية التي يتلقاها الزبون (X_1) وعدد السنوات التي قضاها الزبون في الجامعة (X_2) وقيمة المبيعات للزبون بالدولار حتى تاريخه (Y) ونموذج التنبؤ الذي طُبّق كان:

اكتب النماذج المخفضة لاختبار ما إذا كان أم لا؛

(١) $\beta_1 = \beta_3 = 0$، (٢) $\beta_0 = 0$، (٣) $\beta_3 = 5$، (٤) $\beta_0 = 10$ و(٥) $\beta_1 = \beta_2$،

(٨-٣٣) طُبّق نموذج الانحدار التالي في دراسة لمصادر المياه:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \beta_4 \sqrt{X_{i3}} + \varepsilon_i$$

اكتب النماذج المخفضة لاختبار ما إذا كان أم لا :

(١) $\beta_3 = \beta_4 = 0$ ، (٢) $\beta_3 = 0$ ، (٣) $\beta_1 = \beta_2 = 5$ و(٤) $\beta_4 = 7$

(٨-٣٤) بيّن تكافؤ العبارتين في (8.33) و(8.38) من أجل $r^2_{Y2.1}$

(٨-٣٥) بالإشارة إلى بيانات مثال إنتاجية طاقم عمل في الجدول (٨-٦).

أ - أوجد، من أجل المتغيرات المحولة وفقا لـ (8.41) (١) **X′X** ، (٢) **X′Y** ، (٣) **b** و(٤) $s^2\{\mathbf{b}\}$.

ب - بيّن أن معاملات الانحدار المعيارية التي حصلت عليها في الجزء (أ)(٣) على صلة بمعاملات الانحدار لنموذج الانحدار في متغيراته الأصلية وفقا للعلاقة (8.50).

(٨-٣٦) استنبط العلاقات بين β_k و β'_k في (8.43a) في حالة $p - 1 = 2$.

(٨-٣٧) استنبط العبارة الخاصة بـ **X′Y** في (8.48) في حالة نموذج الانحدار المعياري (8.42) وحيث $p - 1 = 2$.

(٨-٣٨) بالإشارة إلى التمرين (٨-٣٢). اعرض لكل حالة من الحالات الفرضية H_0 مستخدما الصياغة المصفوفية (8.66).

(٨-٣٩) بالإشارة إلى التمرين (٨-٣٣). اعرض لكل حالة من الحالات الفرضية H_0 مستخدما الصياغة المصفوفية (8.66).

(٨-٤٠) (نحتاج إلى حساب التفاضل) استنبط مقدّر المربعات الدنيا تحت النموذج المخفض (8.68).

حيث $\mathbf{C\beta} = \mathbf{h}$ [إرشاد: دالة لاغرانج هي:

$L = (\mathbf{Y} - \mathbf{X\beta})'(\mathbf{Y} - \mathbf{X\beta}) + \lambda'(\mathbf{C\beta} - \mathbf{h})$ حيث: $\lambda' = (\lambda_1, \ldots, \lambda_s)$

(٨ـ٤١) استنبط (8.70)[إرشاد: بيّن أن $SSE(R)\text{-}SSE(F) = (\mathbf{b}_F - \mathbf{b}_R)' \mathbf{X}'\mathbf{X} (\mathbf{b}_F - \mathbf{b}_R)$ وأوجد عبارة $\mathbf{b}_F - \mathbf{b}_R$ من (8.68)].

مشاريع

(٨ـ٤٢) بالإشارة إلى مجموعة البيانات **SMSA** فقد تقرر، للتنبؤ بعدد الأطباء العاملين (Y) في **SMSA**، اعتبار عدد السكان الكلي (X_1) والدخل الشخصي الإجمالي (X_2) كمتغيرين مستقلين. والمسألة الآن هي ما إذا كان وجود متغير مستقل إضافي في النموذج أمرا مفيدا، وإذا كان الأمر كذلك فأي المتغيرات يمكن أن يكون الأكثر فائدة. افترض أن نموذج انحدار متعدد من المرتبة الأولى مناسب.

أ ـ لكل من المتغيرات التالية، احسب معامل التحديد الجزئي علما أن النموذج يتضمن X_1 و X_2 : مساحة المنطقة (X_3)، النسبة المئوية لمن تتجاوز أعمارهم 65 بين السكان (X_4)، عدد أسرّة المستشفيات (X_5) والعدد الإجمالي للجرائم الخطرة (X_6).

ب ـ على أساس نتائج الجزء (أ)، أي المتغيرات المستقلة الأربعة هي الأفضل؟ هل يتجاوز مجموع المربعات الإضافي المرافق لهذا المتغير تلك المرافقة للمتغيرات الثلاثة الأخرى ؟

جـ ـ مستخدما إحصاءة الاختبار F^*، اختبر ما إذا كان المتغير الذي تقرر أنه الأفضل في الجزء (ب) مفيدا في نموذج الانحدار أم لا وذلك عندما يتضمن النموذج X_1 و X_2. استخدم $\alpha = 0.01$. اكتب البديلين، قاعدة القرار، والنتيجة. هل يمكن أن تكون إحصاءة الاختبار F^* للمتغيرات المهمة الثلاثة الأخرى في حجم الإحصاءة F^* هنا؟ ناقش.

(٨ـ٤٣) بالإشارة إلى مجموعة البيانات **SENIC** فقد تقرر، للتنبؤ بمتوسط فترة إقامة المرضى في مستشفى Y، اعتبار العمر (X_1) وخطورة الإصابة (X_2) كمتغيرين مستقلين. والمسألة الآن هي ما إذا كانت إضافة متغير مستقل إلى النموذج

مفيدة أم لا، وإذا كان الأمر كذلك. أي متغير سيكون الأكثر فائدة، افترض أن نموذج انحدار متعدد من المرتبة الأولى مناسب.

أ - لكل من المتغيرات التالية، احسب معامل التحديد الجزئي علما أن النموذج يتضمن X_1 و X_2 : نسبة الزرع الروتينية (X_3)، متوسط الإحصاءات اليومية لعدد المرضى (X_4)، عدد الممرضات (X_5) والخدمات والتسهيلات المتوافرة (X_6)

ب - على أساس النتائج في (أ)، أي المتغيرات الإضافية الأربعة أفضل ؟ هل يتجاوز مجموع المربعات الإضافي المرافق لهذا المتغير تلك المرافقة للمتغيرات الثلاثة الأخرى؟

جـ - مستخدما إحصاءة الاختبار F^* اختبر ما إذا كان المتغير الذي اعتُبر الأفضل في الجزء (ب) مفيدا في نموذج الانحدار أم لا، وذلك عندما يتضمن النموذج X_1 وX_2؛ استخدم $\alpha = 0.05$ اكتب البديلين، وقاعدة القرار، والنتيجة. هل يمكن أن تكون إحصاءة الاختبار F^* للمتغيرات المهمة الثلاثة الأخرى في حجم الإحصاءة F^* هنا؟ ناقش.

(٨-٤٤) بالإشارة إلى التمرين (٧-٢٩) نرغب في اختبار ما إذا كان $\beta_1 = \beta_2$ أم لا. مستخدما طرق المصفوفات، أوجد $SSE(R) - SSE(F)$ وفقا لـ (8.70).

انحدار كثيرات الحدود

ندرس في هذا الفصل نوعا مهما من نماذج الاستجابة المنحنية، ونعني نموذج انحدار كثيرات الحدود. وهو نموذج الاستجابة المنحني الأكثر استخداما في التطبيق العملي نظرا لسهولة معالجته كحالة خاصة من نموذج الانحدار الخطّي العام (7.18) ونناقش أولا بعض نماذج انحدار كثيرات الحدود الشائعة الاستخدام، ثم نقدّم حالتين لتوضيح بعض المشاكل الرئيسة التي تواجهنا في انحدار كثيرات الحدود ونختتم هذا الفصل بمناقشة موجزة لطرائقية سطح الاستجابة.

(٩-١) نماذج انحدار كثيرات الحدود

يمكن أن تتضمن نماذج انحدار كثيرات الحدود واحدا أو اثنين أو أكثر من المتغيرات المستقلة. وفضلا عن ذلك يمكن تقديم كل متغير مرفوعا إلى قوى مختلفة. ونوضّح الآن بعض الإمكانات الرئيسة.

متغير مستقل واحد ـ مرتبة ثانية

يدعى نموذج الانحدار:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \qquad (9.1)$$

حيث $x_i = X_i - \overline{X}$ نموذج مرتبة ثانية بمتغير مستقل واحد لأن المتغير المستقل الوحيد يظهر مرفوعا إلى القوة الأولى والقوة الثانية. لاحظ التعبير عن المتغير المستقل كانحراف عن المتوسط $\overline{X}$ والرمز بـ x_i لانحراف المشاهدة. وسبب استخدام الانحرافات حول المتوسط في نماذج انحدار كثيرات الحدود هو أن X ، X^2 والحدود من قوى أعلى ستكون، في الغالب، مرتبطة ارتباطا عاليا. وكما لاحظنا في الفصل الثامن، يمكن أن يسبب هذا صعوبات حسابية جدّية عند قلب المصفوفة $\mathbf{X'X}$ بغية تقدير معاملات الانحدار. والتعبير عن المتغير المستقل كانحراف عن متوسطه يخفف كثيرا من الخطّية

المتعددة، كما سنوضح في المثال التالي، وينحو إلى تجنب الصعوبات الحسابية.

وكثيرا ماتكتب معاملات الانحدار في انحدار كثيرات الحدود بطريقة مختلفة قليلا، وذلك كي تعكس نمط القوى:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \varepsilon_i \qquad (9.1a)$$

وسنستخدم هذه الرموز في هذا الفصل.

ودالة الاستجابة لنموذج الانحدار ($9.1a$) هي:

$$E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2 \qquad (9.2)$$

وهي دالة قطع مكافىء، وغالبا ما يُقال لها دالة استجابة تربيعية. ويتضمن الشكل (٩-١) مثالين عن دالتي استجابة كثيرة حدود من المرتبة الثانية.

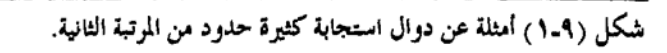
ويمثل معامل الانحدار β_0 متوسط الاستجابة لـ Y عندما $x = 0$، أي عندما $X = \overline{X}$. وغالبا مايدعى معامل الانحدار β_1 معامل التأثير الخطّي بينما يدعى β_{11} معامل تأثير تربيعي.

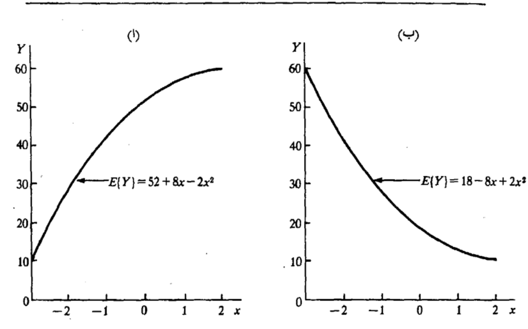
استخدامات نموذج من المرتبة الثانية. لدالة استجابة كثيرة الحدود من المرتبة الثانية (9.2) نوعان أساسيان من الاستخدامات:

١- عندما تكون دالة الاستجابة الصحيحة هي في الحقيقة كثيرة حدود من الدرجة الثانية وتتضمن مركبتي تأثير تجميعيتين خطّية وتربيعية.

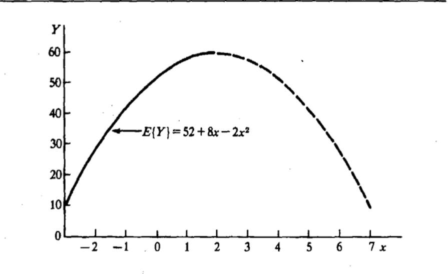
٢- عندما تكون دالة الاستجابة الصحيحة غير معروفة (أو معقدة) ولكن كثيرة حدود من الدرجة الثانية تشكل تقريباً جيدا للدالة الصحيحة.

والنوع الثاني من الاستخدام هو الأكثر شيوعا إلا أنه يؤدي إلى خطورة من نوع خاص عند القيام بعملية تمديد (أو استيفاء خارجي). ولنعتبر ثانية الشكل (٩-١)أ، فقد تشكل دالة الاستجابة هذه توفيقا جيدا تماما للبيانات بين أيدينا. ولكن إذا كانت المعلومات حول $E(Y)$ مطلوبة لقيم أكبر لـ x، فيقود الاستيفاء الخارجي (أوالتمديد) لدالة الاستجابة هذه إلى النتائج المبينة في الشكل (٩-٢)، ونعني تحول دالة الاستجابة


شكل (٩-١) أمثلة عن دوال استجابة كثيرة حدود من المرتبة الثانية.



شكل (٩-٢) تمديد دالة الاستجابة من المرتبة الثانية في الشكل (٩-١)أ



إلى التناقص، مما قد لايتفق مع حقائق الواقع. وتشترك انحدارات كثيرات الحدود بأنواعها جميعا بمثل هذه الخطورة للتمديد (أو الاستيفاء الخارجي)، وبخاصة منها

كثيرات الحدود من مرتبة أعلى. فهي إذ تقدم توفيقات جيدة للبيانات التي بين أيدينا يمكن لها أن تنعطف في اتجاهات غير متوقعة عند التمديد إلى ماوراء البيانات.

متغير مستقل واحد ـ مرتبة ثالثة

ونموذج الانحدار:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_{11} x_i^2 + \beta_{111} x_i^3 + \varepsilon_i \qquad (9.3)$$

حيث $x_i = X_i - \overline{X}$ ، هو نموذج مرتبة ثالثة بمتغير مستقل واحد. ودالة الاستجابة لنموذج الانحدار (9.3) هي:

$$E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2 + \beta_{111} x^3 \qquad (9.4)$$

ويتضمن الشكل (٩-٣) مثالين لدالتي استجابة كثيرة حدود من المرتبة الثالثة.

متغير مستقل واحد ـ مراتب أعلى

في الغالب لا تُستخدم نماذج كثيرات الحدود بمتغير مستقل قوته أعلى من ثلاثة، إذ يصبح تفسير معاملات الانحدار صعبا في نماذج كهذه، وقد تؤدي إلى أخطاء مرتفعة في الاستيفاءات الداخلية، وفي الاستيفاءات الخارجية حتى الصغير منها. وفي هذا المجال لابد من معرفة أنه يمكن دائما إيجاد نموذج كثيرة حدود من مرتبة عالية بصورة كافية بحيث تشكّل توفيقا تاما للبيانات. وعلى سبيل المثال، فإن توفيق دالة انحدار كثيرة حدود بمتغير مستقل واحد من المرتبة $n - 1$ يستمر عبر الـ n من قيم Y الملحوظة جميعا. وبالتالي ينبغي للمرء أن يكون حذرا من استخدام كثيرات حدود من مرتبة عالية لمجرد الحصول على توفيق جيد. فقد لاتبين دوال انحدار كهذه بوضوح العناصر الأساسية لعلاقة الانحدار بين X و Y كما أنها قد تقود إلى استيفاءات داخلية وخارجية خاطئة.

متغيران مستقلان ـ مرتبة ثانية

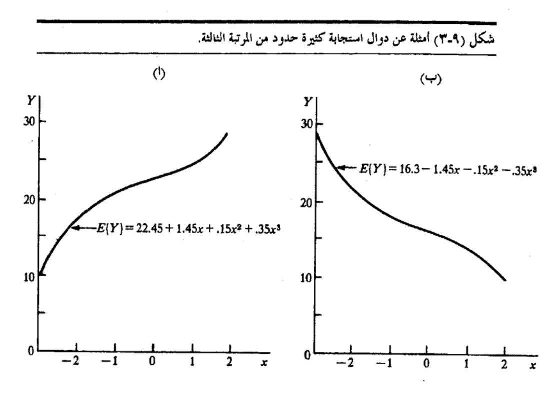
نموذج الانحدار:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i \qquad (9.5)$$

حيث:

$$x_{i1} = X_{i1} - \overline{X}_1$$
$$x_{i2} = X_{i2} - \overline{X}_2$$

شكل (٩-٣) أمثلة عن دوال استجابة كثيرة حدود من المرتبة الثالثة.



هو نموذج مرتبة ثانية بمتغيرين مستقلين. ودالة الاستجابة هي:

$$E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 \qquad (9.6)$$

وهي معادلة مقطع مخروطي. لاحظ أن نموذج الانحدار (9.5) يتضمن مركّبات خطّية وتربيعية منفصلة لكل من المتغيرين المستقلين وحَدا جدائيا. ويمثل الحد الجدائي تأثيرات التفاعل بين x_1 و x_2 كما وردت في الفصل السابع. ويدعى المعامل β_{12} ، في الغالب، معامل تأثير التفاعل.

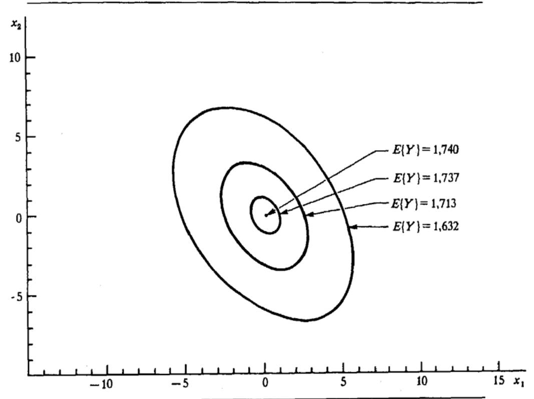
وتمثل دالة الاستجابة من المرتبة الثانية (9.6) بمتغيرين مستقلين النوعين الأساسيين من السطوح الموضّحة في الشكل (٧-٣). وتشكل الحوافّ المستقرة والصاعدة حالات حدّية لهذين النوعين الأساسيين من سطوح الاستجابة.

ويكون من الأيسر، في الغالب، تصوير سطح الاستجابة من المرتبة الثانية (9.6) بدلالة منحنيات خطوط التساوي. ويتضمن الشكل (٩-٤) تمثيلا بدلالة منحنيات خطوط التساوي لدالة الاستجابة:

$$E\{Y\} = 1{,}740 - 4x_1^2 - 3x_2^2 - 3x_1x_2 \qquad (9.7)$$

لاحظ أن لسطح الاستجابة هذا قيمة عظمى عند $x_1 = 0$ و $x_2 = 0$.

شكل (٩-٤) مثال عن سطح استجابة تربيعي بدلالة منحنيات خطوط التساوي:
$E\{Y\} = 1{,}740 - 4x_1^2 - 3x_2^2 - 3x_1x_2$



ويمكن تكييف نماذج كثيرات حدود بمتغيرين مستقلين (أو أكثر) بصورة جيدة إلى حالات تكون دالة الاستجابة فيها غير معروفة والمطلوب تطوير نموذج مناسب بصورة تجريبية.

ملاحظة

يعتبر الحد الجدائي $\beta_{12}\, x_1x_2$ في (9.6) حداً من المرتبة الثانية، مثله مثل $\beta_{11}\, x_1^2$ أو $\beta_{22}\, x_2^2$ ، ويمكن رؤية السبب بسهولة لدى كتابة الحدين الأخيرين، على الترتيب، في صيغة $\beta_{11}x_1x_1$ و $\beta_{22}x_2x_2$.

ثلاثة متغيرات مستقلة - مرتبة ثانية

نموذج الانحدار من المرتبة الثانية بثلاثة متغيرات مستقلة هو:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{33} x_{i3}^2 + \beta_{12} x_{i1} x_{i2} + \beta_{13} x_{i1} x_{i3} + \beta_{23} x_{i2} x_{i3} + \varepsilon_i \quad (9.8)$$

حيث:

$$x_{i1} = X_{i1} - \overline{X}_1$$

$$x_{i2} = X_{i2} - \overline{X}_2$$

$$x_{i3} = X_{i3} - \overline{X}_3$$

ودالة الاستجابة لنموذج الانحدار هذا هي:

$$E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \quad (9.9)$$

والمعاملات β_{12}، β_{13} و β_{23} هي معاملات تأثيرات التفاعل للتفاعلات بين أزواج من المتغيرات المستقلة.

استخدامات نماذج انحدار كثيرات الحدود

لايشكّل توفيق نماذج انحدار كثيرات الحدود مسألة جديدة باعتبارها، كما رأينا في الفصل السابع، حالات خاصة من نموذج الانحدار الخطي العام (7.18). وبالتالي تنطابق هنا جميع النتائج السابقة الخاصة بتوفيق نموذج، وكذلك النتائج السابقة حول القيام باستقراءات.

وعند استخدام نموذج انحدار كثيرة حدود كتقريب لدالة الانحدار الصحيحة، فسيقوم المرء في الغالب، بتوفيق نموذج مرتبة ثانية أو مرتبة ثالثة ثم يستطلع ما إذا كان نموذج من مرتبة أدنى مناسبا. وعلى سبيل المثال، يمكن توفيق النموذج:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \beta_{111} x_i^3 + \varepsilon_i$$

بمتغير مستقل واحد، مع الأمل بأنه يمكن إلغاء الحد التكعيبي وربما الحد التربيعي أيضا. وهكذا، قد نرغب في اختبار ما إذا كان $\beta_{111} = 0$ أم لا، أو ما إذا كان كل من $\beta_{11} = 0$ و $\beta_{111} = 0$ أم لا، وفي الغالب يجري القيام باختبارات مماثلة في نماذج انحدار كثيرات حدود بمتغيرين مستقلين أو أكثر.

(٩-٢) مثال ١- متغير مستقل واحد

ونوضح الآن بعض الأنـواع الرئيسـة لتحليـلات تتـم عـادة في نمـاذج انحـدار كثـيرات الحدود بمتغير مستقل واحد.

عرض المسألة

رغب مُحلّل من الهيئة المسؤولة عن سلسلة من محـلات الكافتيريـا تحـرّي العلاقـة بين عدد آلات صرف القهوة بطريقـة الخدمـة الذاتيـة في محـل كافتيريـا وبـين مبيعـات القهوة. وقد اختير للتجربة أربعـة عشـر محـلا تتشـابه مـن حيـث حجـم العمـل، نـوع الزبائن، والموقع. وتراوح عدد آلات الخدمة الذاتيـة المتوافـرة في المحـلات الـتي تناولتهـا التجربة بين صفر (يقدم القهوة هنا عامل) وبـين سـت آلات، وقـد خصصـت الآلات إلى المحلات بطريقة عشوائية.

ويتضمـن الجـدول (٩-١) نتـائج الدراسـة التجريبيـة. وقيسـت المبيعـات بمئـات الجالونات من القهوة المباعة.

توفيق نموذج

يعتقد المحلل أن العلاقة بين المبيعات وعدد آلات الخدمة الذاتية هي علاقة تربيعيـة ضمن مدى المشاهدات، إذ ينبغي أن تزيد المبيعات مع زيـادة عـدد الآلات، ولكـن إذا أصبح المكان مزدحما بالآلات تبدأ الزيـادة بالـتراجع. وبالتـالي يرغـب المحلـل في توفيـق نموذج الانحدار التربيعي:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \varepsilon_i \qquad (9.10)$$

حيث : $x_i = X_i - \overline{X}$

ويتوقع المحلل، فضـلا عـن ذلـك، أن تتبـع حـدود الخطـأ ε_i، بصـورة مقبولـة، التوزيـع الطبيعي بتباين ثابت.

والمصفوفتـان **X** و **Y** لهـذا التطبيـق معطيتـان في الجـدول (٩-٢). لاحـظ أن المصفوفة **X** تتضمن عمودا مـن المقـادير 1 وعمـودا مـن مشـاهدات المتغـير المسـتقل x (معبرا عنها كانحرافات عن متوسطها $\overline{X}=3$)، وعمودا من قيم x^2.

جدول (٩-١) بيانات مثال مبيعات القهوة في كافتيريا

الكافيتريا i	عدد آلات الخدمة الذاتية X_i	مبيعات القهوة (بمئات الجالونات) Y_i
1	0	508.1
2	0	498.4
3	1	568.2
4	1	577.3
5	2	651.7
6	2	657.0
7	3	713.4
8	3	697.5
9	4	755.3
10	4	758.9
11	5	787.6
12	5	792.1
13	6	841.4
14	6	831.8

لاحظ أيضا في الجدول (٩-٢) أن كلا من القيم الصغيرة والكبيرة لـ x يقابلها قيم كبيرة لـ x^2. وكنتيجة لذلك، وبسبب تناظر الانحرافات، فإن معامل الارتباط البسيط بين X و X^2 في الجدول (٩-٢)، هو $r = 0$ ولو أن نموذج انحدار كثيرات الحدود استخدم قيم X الأصلية:

X :	0	0	1	1	...	6	6
X^2 :	0	0	1	1	...	36	36

فستوافق القيم الصغيرة لـ X قيما صغيرة فقط لـ X^2 ، وسيوافق القيم الكبيرة لـ X قيما كبيرة فقط لـ X^2. وفي هذه الحالة، فإن معامل الارتباط البسيط بين X و X^2 سيصبح $r = 0.961$. وهذا يوضّح أن استخدام انحرافات قيم المتغير المستقل عن المتوسط يقود إلى ارتباط أقل بكثير بين المتغيرات X في نماذج انحدار كثيرات الحدود. وفي مثالنا هنا ، نجد أن x و x^2 في الحقيقة، غير مرتبطين بسبب تناظر الانحرافات x.

ومن هنا فصاعدا فإن الحسابات هي مجرد روتين. ويمكن القيام بالحسابات المصفوفية بصورة يدوية، كما أوضحنا في الفصل السابع، أو نستخدم برنامج حاسوب للانحدار المتعدد. وبما أننا لانواجه هنا أية مشاكل جديدة. فنقدم في الجدول (٩-٣) ببساطة النتائج الأساسية لبرنامج حاسوب، بما في ذلك مجاميع المربعات الإضافية التي نحتاجها ومصفوفة الـ $\mathbf{s}^2\{\mathbf{b}\}$.

جدول (٩-٢) مصفوفات البيانات لمثال مبيعات القهوة في كافيتيريا

$$
\mathbf{Y}=\begin{bmatrix}508.1\\498.4\\568.2\\577.3\\651.7\\657.0\\713.4\\697.5\\755.3\\758.9\\787.6\\792.1\\841.4\\831.8\end{bmatrix}
\qquad
\mathbf{X}=\begin{bmatrix}1&-3&9\\1&-3&9\\1&-2&4\\1&-2&4\\1&-1&1\\1&-1&1\\1&0&0\\1&0&0\\1&1&1\\1&1&1\\1&2&4\\1&2&4\\1&3&9\\1&3&9\end{bmatrix}
$$

(الأعمدة في X: x ، x^2)

جدول (٩-٣) نتائج الانحدار لمثال مبيعات القهوة في كافيتريا

(أ) معاملات الانحدار

معامل الانحدار	معامل الانحدار المقدّر	الانحدار المعياري المقدر	t^*
β_0	705.474	3.208	219.91
β_1	54.893	1.050	52.28
β_{11}	-4.249	0.606	-7.01

(ب) تحليل التباين

مصدر التغير	SS	df	MS
الانحدار	171,773	2	85,887
x	168,741	1	168,741
$x^2 \mid x$	3,033	1	3,033
الخطأ	679	11	61.7
المجموع	172,453	13	

(جـ) مصفوفة $s^2\{b\}$

$$
\begin{bmatrix}10.2912&0&-1.4702\\0&1.1026&0\\-1.4702&0&0.3675\end{bmatrix}
$$

ودالة الانحدار التوفيقية هي:

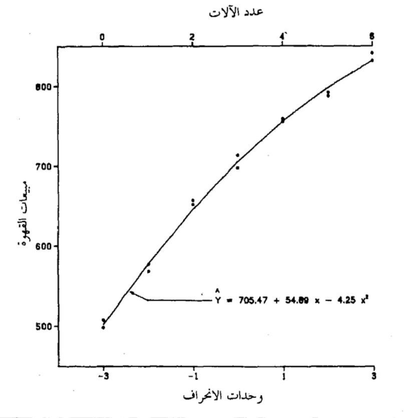
$$\hat{Y} = 705.47 + 54.89x - 4.25x^2 \qquad (9.11)$$

وقد رُسمت دالة الاستجابة هذه في الشكل (٩-٥)، بالإضافة إلى البيانات الأصلية، ونبين السلم الأفقي معبرا عنه في أسفل الشكل بوحدات الانحرافات x وفي أعلى الشكل بالوحدات الأصلية X.

الشكل الجبري للمعادلات الناظمية. الشكل الجبري لمعادلات المربعات الدنيا الناظمية:

$$\mathbf{X'X\,b = X'Y}$$

شكل (٩-٥) انحدار كثيرة حدود توفيقية من المرتبة الثانية - مثال مبيعات القهوة في كافتيريا.



لنموذج انحدار كثيرات الحدود من المرتبة الثانية (9.10)، يمكن الحصول عليها بسهولة من (7.68) باستبدال x_i و x_i^2 بـ X_{i1} و X_{i2} على الترتيب. وبما أن $\sum x_i = 0$ ، فهذا

يُنتج المعادلات الناظمية التالية:

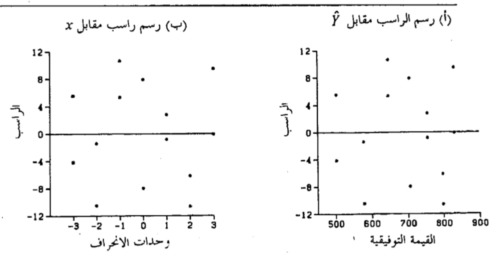
$$\begin{aligned} \sum Y_i &= nb_0 + b_{11}\sum x_i^2 \\ \sum x_i Y_i &= b_1 \sum x_i^2 + b_{11}\sum x_i^3 \\ \sum x_i^2 Y_i &= b_0 \sum x_i^2 + b_1 \sum x_i^3 + b_{11}\sum x_i^4 \end{aligned} \tag{9.12}$$

تحليل صلاحية النموذج

تحليل الراسب. ولدراسة صلاحية نموذج الانحدار (9.10) للبيانات المعطاة فقد رسم المحلل الرواسب e_i مقابل القيم التوفيقية، كما هو مبين في الشكل (٩-٦)أ، وكذلك مقابل المتغير المستقل x_i معبرا عنه بوحدات انحراف، كما هو مبين في الشكل (٩-٦)ب. ولانقدم حسابات e_i باعتبارها مجرد روتين.

ولاتوجد، بين الرواسب، انحرافات نمطية واضحة عن الصفر، وذلك عندما يزداد أي من $\hat{Y}$ أو x، مما يقترح جودة توفيق دالة الاستجابة التربيعية. ويؤدي رسم الانتشار في الشكل (٩-٥) إلى هذه النتيجة أيضا. وفضلا عن ذلك لاتوجد، في الشكلين (٩-٦)أ و(٩-٦)ب ، أية نزعة للتغير بصورة نمطية في انتشار الرواسب، وهكذا يبدو افتراض ثبات تباين الخطأ افتراضا معقولا. ولايقدم رسم الاحتمال الطبيعي (غير مبين هنا) أية دلالة قوية لابتعاد توزيع حدود الخطأ عن الطبيعي.

شكل (٩-٦) رسوم الراسب لمثال مبيعات القهوة في كافيتريا.



واستنادا إلى هذه الدراسة حول صلاحية النموذج كان المحلل مستعدا بكل ترحاب لاستنتاج أن نموذج انحدار الخطأ الطبيعي (9.10) مع تباين خطأ ثابت هو نموذج مناسب هنا.

اختبار من أجل دالة استجابة تربيعية. بما أن لدينا مشاهدتين مكررتين عند كل مستوى لـ x ، فكان يمكن للمحلل أن يستخدم الاختبار الرسمي (7.57) لصلاحية دالة الانحدار والبدائل هنا هي:

$$H_0: E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2$$
$$H_a: E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2 \qquad (9.13)$$

وقد عرضنا سابقا في الجدول (٩-٣)ب نتائج التحاين الأساسية ومن بيانات الجدول (٩-٢) نحصل على مجموع مربعات الخطأ البحت (4.11) كما يلي:

$$SSPE = (508.1 - 503.25)^2 + (498.4 - 503.25)^2 + (568.2 - 572.75)^2 + \ldots + (831.8 - 836.6)^2 = 292$$

لاحظ أن $\overline{Y}_1 = 503.25$ من أجل $x = -3$، $\overline{Y}_2 = 572.75$ من أجل $x = -2$، وهلمجرا.

وهناك $c = 7$ مستويات متميزة لـ x وهكذا يكون $n - c = 14 - 7 = 7$ درجات حرية موافقة لـ $SSPE$ ، وبالتالي لدينا:

$$MSPE = \frac{SSPE}{7} = \frac{292}{7} = 41.7$$

وباستخدام (4.19) نحن الآن في موقع يمكننا معه الحصول على مجموع مربعات نقص التوفيق:

$$SSLF = SSE - SSPE = 679 - 292 = 387$$

وتوجد $c - p = 7 - 3 = 4$ درجات حرية موافقة لـ $SSLF$. (لنتذكر أننا مضطرون لتقدير $p = 3$ معالم في دالة الانحدار التوفيقية) وهكذا نجد:

$$MSLF = \frac{SSLF}{4} = \frac{387}{4} = 96.8$$

وتصبح الإحصاءة (7.57b) هنا:

$$F^* = \frac{MSLF}{MSPE} = \frac{96.8}{41.7} = 2.32$$

وبافتراض مستوى معنوية 0.05، نحتاج إلى $F(.95;4,7)=4.12$. وبما أن $F^* = 2.32 \leq 4.12$ فنستنتج H_0، أي أن دالة الانحدار التربيعية مناسبة.

اختبار ما إذا كان β_{11} مساويا للصفر

اختبار t. درس المحلل بعد ذلك ما إذا كان يمكن إسقاط الحد التربيعي من النموذج. ولذلك رغب في اختبار الفرضية.

$$H_0 : \beta_{11} = 0 \qquad (9.14)$$
$$H_a : \beta_{11} \neq 0$$

وتتضمن H_0 أنه لايوجد تأثير تربيعي في دالة الاستجابة.

ويشير الجدول (٩-٣)أ إلى أن:

$$t^* = \frac{b_{11}}{s\{b_{11}\}} = \frac{-4.249}{0.606} = -7.012$$

ومن أجل مستوى معنوية 0.05 نحتاج إلى $t(.975\,;\,11) = 2.201$. وقاعدة القرار هي:

إذا كان $|t^*| \leq 2.201$ استنتج H_0

إذا كان $|t^*| > 2.201$ استنتج H_a

وبما أن $|t^*| = 7.012 > 2.201$ نستنتج H_a، أي أن التأثير التربيعي موجود وبالتالي ينبغي الاحتفاظ بالحد التربيعي في النموذج.

اختبار F جزئي. كان يمكن للمحلل أن يستخدم اختبار F الجزئي لاختيار النتيجة المناسبة في (9.14). وفي الحقيقة، فقد حدد ترتيب دخول المتغيرين x و x^2 إلى برنامج توفيق الانحدار بحيث يحصل على مجاميع المربعات $SSR(x)$ و $SSR(x^2 | x)$ في مخرجات الحاسوب. وبالاستفادة من إحصاءة اختبار F الجزئي (8.22) والنتيجة في الجدول (٩-٣)ب، نحصل على:

$$F^* = \frac{MSR(x^2 | x)}{MSE} = \frac{3,033}{61.7} = 49.2$$

ولمستوى معنوية 5 بالمائة، نحتاج إلى $F(.95\,;\,1,11) = 4.84$. وبما أن $F^* = 49.2 > 4.84$ فهذا يقودنا إلى استنتاج H_a، كما هو الحال مع الاختبار t.

ملاحظة

يمكن أن نلاحظ هنا العلاقة التي ناقشناها في الفصل السابق بين الاختبارين t وF لفرضية أن معامل انحدار يساوي الصفر. ولدينا من أجل إحصاءتي الاختبار:

$$(t^*)^2 = (-7.012)^2 = 49.2 = F^*$$

تقدير معاملات انحدار

رغب المحلل بعد ذلك في الحصول على حدود ثقة لمعاملي الانحدار β_1 و β_{11} بمعامل ثقة عائلي 0.90 باستخدام طريقة بونفرّوني.

نرغب هنا بعبارتين وبالتالي لدينا من (7.47a):

$$B = t[1 - 0.10 / 2(2);\ 11] = t(.975;11) = 2.201$$

ونجد من الجدول (٩-٣)أ:

$$b_1 = 54.893 \qquad s\{b_1\} = 1.050$$
$$b_{11} = -4.249 \qquad s\{b_{11}\} = 0.606$$

ومن (7.47) تكون حدود ثقة بونفرّوني بالتالي $54.893 \pm 2.201(1.050)$ و $-4.249 \pm 2.201(0.606)$، مما يعطي فترتي الثقة:

$$52.58 \le \beta_1 \le 57.20$$
$$-5.58 \le \beta_{11} \le -2.92$$

وكان المحلل مرتاحا لدقة هاتين العبارتين، وهو يشعر بأن الفترتين ضيقتان ضيقا كافيا لإمداده بمعلومات متزامنة موثوقة عن الحجم النسبي للتأثيرين الخطّي والتربيعي.

معامل التحديد المتعدّد

وللحصول على مقياس وصفي لدرجة الصلة بين مبيعات القهوة وعدد آلات الخدمة الذاتية، قام المحلل بحساب معامل التحديد المتعدّد المعطى في (7.35) مستخدما بيانات الجدول (٩-٣)ب، فوجد:

$$R^2 = \frac{SSR}{SSTO} = \frac{171,773}{172,453} = 0.996$$

ويبين هذا المقياس أن التغير في مبيعات القهوة ينخفض بنسبة 99.6 بالمائة عند استخدام علاقة تربيعية في عدد آلات الخدمة الذاتية.

نلاحظ أن معامل التحديد المتعـدّد R^2 هـو المقيـاس المناسـب هنـا وليـس معـامل التحديد البسيط r^2 باعتبار أن النموذج (9.10) هو نموذج انحدار متعدد بالرغم من أنـه يتضمن متغيرا مستقلا واحدا فقط. وفي الانحدار المنحني يدعى معامل الارتبـاط المتعـدد R، أحيانا، دليل الارتباط.

تقدير متوسط استجابة

كان المحلل مهتما، على وجه الخصـوص، بمتوسـط الاستجابة في حالـة $X_h = 3$ آلات خدمة ذاتية. وقد رغب بتقدير متوسط الاسـتجابة هـذا بمعـامل ثقـة 98 بالمائـة. وتقدير الفترة المناسب معطى في (7.54). وفي مثالنـا، حيـث $x_h = X_h - \overline{X} = 3-3=0$ لدينا:

$$\boldsymbol{X}_h = \begin{bmatrix} 1 \\ x_h \\ x_h^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

ومتوسط الاستجابة المقدَّر $\hat{Y}_h$ الموافق لـ $\mathbf{X}_h$ استنادا إلى (7.50) هو:

$$\hat{Y}_h = \mathbf{X}'_h\mathbf{b} = [1 \quad 0 \quad 0]\begin{bmatrix} 705.474 \\ 54.893 \\ -4.249 \end{bmatrix} = 705.474$$

وباستخدام نتائج الجدول (٩-٣)جـ الخاصة بـ $\mathbf{s}^2\{\mathbf{b}\}$، نحصل الآن، عند التعويض في (7.53) على:

$$s^2\{\hat{Y}_h\} = \mathbf{X}'_h\ \mathbf{s}^2\{\mathbf{b}\}\ \mathbf{X}_h$$

$$= [1 \ 0 \ 0]\begin{bmatrix} 10.2912 & 0 & -1.4702 \\ 0 & 1.1026 & 0 \\ -1.4702 & 0 & 0.3675 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 10.2912$$

أو $s\{\hat{Y}_h\} = 3.208$. ونحتـاج إلى $t(.99;\ 11) = 2.718$. وبالتـالي يكـون حـدا الثقـة، $705.474 \pm 2.718(3.208)$ ونحصل على فترة الثقة:

$$696.8 \leq E\{Y_h\} \leq 714.2$$

بمعامل ثقة 0.98 ، ويمكن للمحلل أن يستنتج أن متوسط مبيعات القهوة عند استخدام ثلاث آلات خدمة ذاتية يتراوح بين 696.8 و714.2 من مئات الجالونات.

دالة انحدار بدلالة X

ولأغراض كتابة تقرير، يرغب المحلل في التعبير عن دالة الانحدار التوفيقية (9.11) بدلالة X بدلا من أن تكون بدلالة الانحرافات $x = X - \bar{X}$. ودالة الانحدار التوفيقية المكافئة بدلالة المتغير الأصلي X هي:

$$\hat{Y} = b_0' + b_1' X + b_{11}' X^2 \qquad (9.15)$$

حيث نحصل على المعاملات كما يلي:

$$b_0' = b_0 - b_1 \bar{X} + b_{11} \bar{X}^2 \qquad (9.15a)$$

$$b_1' = b_1 - 2b_{11} \bar{X} \qquad (9.15b)$$

$$b_{11}' = b_{11} \qquad (9.15c)$$

وفي مثالنا، حيث $\bar{X} = 3$ نجد:

$b_0' = 705.474 - 54.893(3) + (-4.249)(3)2 = 502.554$

$b_1' = 54.893 - 2(-4.249)(3) = 80.387$

$b_{11}' = - 4.249$

وهكذا تكون دالة الانحدار التوفيقية بدلالة X كما يلي:

$$\hat{Y} = 502.554 + 80.387X - 4.249X^2$$

والقيم التوفيقية والرواسب الناتجة عن دالة الانحدار بدلالة X تبقى نفسها بالضبط كما في دالة الانحدار بدلالة الانحرافات x. وكما نوّهنا سابقا، فإن سبب استخدام النموذج (9.10)، المعبّر عنه بدلالة الانحرافات x ، هو تجنب الصعوبات الحسابية الكبيرة التي تسببها الخطّية المتعددة بين X و X^2 ، المتأصلة في انحدار كثيرات الحدود.

ملاحظة

لاتنطبق الانحرافات المعيارية المقدّرة لمعاملات الانحدار في الجدول (٩ـــ٣)أ على معاملات الانحدار بدلالة X في (9.15). وإذا رغبنا بالانحرافات المعيارية المقدّرة لمعاملات الانحدار بدلالة X فيمكن الحصول عليها من $\mathbf{s^2\{b\}}$ في الجدول (٩–٣)جـ مستخدمين النظرية (6.47) حيث نستنبط مصفوفة التحويل **A** من (9.15 a,b,c).

(٩-٣) مثال ٢- متغيران مستقلان

سنناقش الآن مثالا آخر لانحدار كثيرات الحدود يتضمن متغيرين مستقلين. وبدلا من المضي في هذا المثال عبر جميع المراحل المختلفة للتحليل، كما فعلنا في المثال الأول، سنركّز هنا بصورة رئيسة على تأثيرات التفاعل والتأثيرات التربيعية.

جدول (٩-٤) بيانات مثال خلايا الطاقة.

خلية	(١) معدّل الشحن	(٢) درجة الحرارة	(٣) قيم مرمّزة	(٤)	(٥) عدد الدورات
i	X_{i1}	X_{i2}	x_{i1}	x_{i2}	Y_i
1	0.6	10	-1	-1	150
2	1.0	10	0	-1	86
3	1.4	10	1	-1	49
4	0.6	20	-1	0	288
5	1.0	20	0	0	157
6	1.0	20	0	0	131
7	1.0	20	0	0	184
8	1.4	20	1	0	109
9	0.6	30	-1	1	279
10	1.0	30	0	1	235
11	1.4	30	1	1	224
	$\overline{X}_1 = 1.0$	$\overline{X}_2 = 20$			

المصدر: S.M. Sidik, H. F. Leibecki, and J.M. Bozek, *Cycles Till Failure of Silver-Zinc Cells with Competing Failure Modes - Preliminary Data Analysis*, NASA Technical Memorandum 81556, 1980.

عرض المسألة

درس باحث تأثيرات معدل الشحن الكهربائي ودرجة الحرارة على حياة نوع جديد من خلايا الطاقة. وقد نُفّذت التجربة بحيث ضُبطت معدّلات الشحن (X_1) عند ثلاثة مستويات (0.60، 1.0 و1.4 أمبير) وضبطت درجة حرارة المحيط (X_2) عند ثلاثة مستويات (10، 20 و30 درجة مئوية). وقد أُبقيت العوامل المتعلقة بتفريغ خلية الطاقة عند مستويات ثابتة. وقد قيس عمر خلية الطاقة (Y) بدلالة عدد دورات التفريغ – الشحن التي عانتها الخلية قبل فشلها النهائي. والبيانات الناتجة عن الدراسة مبينة في

الأعمدة (١)، (٢) و(٥) من الجدول (٩-٤).

لم يكن الباحث متأكدا من طبيعة دالة الاستجابة ضمن مدى العوامل المدروسة. وبالتالي قرر توفيق نموذج انحدار كثيرة حدود من المرتبة الثانية (9. 5):

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2 + \beta_{12} x_{i1} x_{i2} + \varepsilon_i \qquad (9.16)$$

ودالة الاستجابة لهذا النموذج هي:

$$E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 \qquad (9.16a)$$

وبسبب الطبيعة المتوازنة لمستويات X_1 و X_2 المدروسة، فقد عرَّف الباحث انحرافات X_1 و X_2 عن متوسطاتها الموافقة بوحدات مريحة حسابيا، وذلك كما يلي:

$$x_{i1} = \frac{X_{i1} - \overline{X}_1}{0.4} = \frac{X_{i1} - 1.0}{0.4}$$
$$x_{i2} = \frac{X_{i2} - \overline{X}_2}{10} = \frac{X_{i2} - 20}{10} \qquad (9.16b)$$

حيث يمثل المقام لكل متغير الفرق بين مستويين متجاورين. ومتغيرات الانحراف هذه مبينة في العمودين (٣) و (٤) من الجدول (٩-٤). لاحظ أن الانحرافات المعرّفة في (9.16b) تقود إلى ترميز بسيط لمستويات X_1 و X_2 بدلالة 1- ، 0 و1.

وكان الباحث مهتما، على وجه الخصوص، بما إذا كان ينبغي للنموذج أن يتضمن تأثيرات تفاعل وتأثيرات منحنية وذلك ضمن المدى المدروس للمتغيرات X.

تطوير نموذج

يتضمن الجدول (٩-٥)أ نتائج الانحدار الأساسية لتوفيق النموذج (9.16). وقد تحرّى الباحث أولا صلاحية نموذج الانحدار هذا للبيانات المتوافرة. ورسومات الرواسب مقابل $\hat{Y}$ ، x_1 و x_2 مبينة في الشكل (٩-٧)، وكذلك رسم احتمال طبيعي. ولايقترح أي من هذه الرسوم أي قدر كبير من عدم ملاءمة نموذج الانحدار (9.16) ومعامل الارتباط بين الرواسب المرتبة وتوقعاتها تحت الطبيعية هو 0.974 ، مما يدعم افتراض طبيعية حدود الخطأ (انظر الجدول ٤-٣).

ويمكن الحصول على مؤشر آخر لصلاحية النموذج (9.16) من الاختبار الرسمي

(7.57) لجودة توفيق دالة الانحدار (9.16a)، باعتبار أنه تتوافر عدة تطبيقات عند $x_1=0$ و$x_2 = 0$ ومجموع مربعات الخطأ البحت (4.11) بسيط هنا، إذ لا تتوافر التكرارات إلا في تركيب واحد فقط من تراكيب المستويات:

$$SSPE = (157 - 157.33)^2 + (131 - 157.33)^2 + (184 - 157.33)^2 = 1,404.67$$

جدول (٩-٥) نتائج الانحدار لنموذج كثيرة الحدود من المرتبة الثانية (9.16) مثال خلايا الطاقة.

(أ) معامل الانحدار

معامل الانحدار	معامل الانحدار المقدّر	الانحراف المعياري المقدّر	t^*
β_0	162.84	16.61	9.81
β_1	-55.83	13.22	-4.22
β_2	75.50	13.22	5.71
β_{11}	27.39	20.34	1.35
β_{22}	-10.61	20.34	-0.52
β_{12}	11.50	16.19	0.71

(ب) تحليل التباين

مصدر التغير	SS	df	MS
الانحدار	55,366	5	11,073
x_1	18,704	1	18,704
$x_2 \mid x_1$	34,201	1	34,201
$x_1^2 \mid x_1, x_2$	1,646	1	1,646
$x_2^2 \mid x_1, x_2, x_1^2$	285	1	285
$x_1x_2 \mid x_1, x_2, x_1^2, x_2^2$	529	1	529
الخطأ	5,240	5	1,048
المجموع	60,606	1	

وبما أنه يوجد هنا $c = 9$ تراكيب متميزة لمستويات المتغيرات فيتوافق مع $SSPE$ هنا $n - c = 11 - 9 = 2$ درجة حرية. وفضلا عن ذلك، فإن $SSE = 5240$ وفقا للجدول (٩-٥)ب، وبالتالي يكون مجموع مربعات نقص التوفيق (4.19):

$$SSLF = SSE - SSPE = 5,240 - 1,404.67 = 3,835.33$$

ويوافقها ثلاث درجات حرية $p = 9 - 6 = 3$ - c. وهكذا تكون إحصاءة الاختبار

(7.57b) لاختبار ملاءمة دالة الانحدار (9.16a) هي:

$$F^* = \frac{SSLF}{c-p} \div \frac{SSPE}{n-c} = \frac{3,835.33}{3} \div \frac{1,404.67}{2} = 1.82$$

شكل (٩-٧) رسومات الراسب لمثال خلايا الطاقة.

(أ) رسم الراسب مقابل $\hat{Y}$

(ب) رسم الراسب مقابل x_1

(جـ) رسم الراسب مقابل x_2

(د) رسم الاحتمال الطبيعي

القيمة المتوقعة

ومن أجل $\alpha = 0.05$ نحتاج إلى $F(.95;3,2) = 19.2$. وبما أن $F^* = 1.82 \leq 19.2$. فنستنتج وفقا لقاعدة القرار (7.57c) أن دالـة انحـدار كثيرة الحـدود مـن المرتبـة الثانيـة (9.16a) ملائمة.

وقد التفت الباحث، الآن إلى دراسة ما إذا كان نموذج المرتبة الأولى ملائما. وبدائل الاختبار هي:

$$H_0 : \beta_{11} = \beta_{22} = \beta_{12} = 0$$

H_a : ليست كل المعاملات β في H_0 مساوية للصفر

وإحصاءة اختبار F الجزئي (8.25) هي هنا:

$$F^* = \frac{SSR(x_1^2, x_2^2, x_1x_2 | x_1x_2)}{3} \div MSE$$

ووفقا لتوقعـات هـذا الاختبـار فقـد أدخـل البـاحث المتغيرات X في برنـامج الانحـدار الحاسوبي بالترتيب x_1، x_2، x_1^2، x_2^2 و x_1x_2 ونبين في الجدول (٩-٥)ب جـدول تحليـل التباين، بما في ذلك مجاميع المربعـات الإضافيـة و بالتـالي يمكـن الحصـول علـى مجمـوع المربعات الإضافي المطلوب كما يلي:

$$\begin{aligned} SSR(x_1^2, x_2^2, x_1x_2 | x_1x_2) &= SSR(x_1^2 | x_1, x_2) + SSR(x_2^2 | x_1, x_2, x_1^2) \\ &\quad + SSR(x_1x_2 | x_1, x_2, x_1^2, x_2^2) \\ &= 1{,}646 + 285 + 529 = 2{,}460 \end{aligned}$$

وإحصاءة الاختبار هي:

$$F^* = \frac{2{,}460}{3} \div 1{,}048 = 0.78$$

ولمستوى معنوية $\alpha = 0.05$ نحتاج إلى $F(0.95; 3,5) = 5.41$ وبما أن $F^* = 0.78 \leq 5.41$ فنستنتج H_0، أي عدم وجود تأثـيرات تفـاعل منحنيـة، وبالتـالي يكـون نمـوذج المرتبـة الأولى ملائما لمدى معدلات الشحن ودرجات الحرارة المدروسة.

وهكذا فقد قرر الباحث استخدام نموذج المرتبة الأولى:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \tag{9.17}$$

وحصل على دالة الاستجابة التوفيقية:

$$\hat{Y} = 172.00 - 55.83x_1 + 75.50x_2 \tag{9.18}$$

ونلاحظ أن معاملي الانحدار b_1 و b_2 هما نفساهما كمـا في الجـدول (٩ـ٥)أ الخـاص

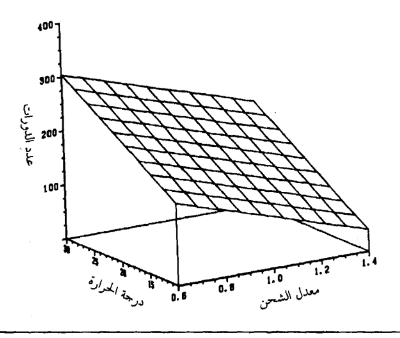
بالنموذج التوفيقي من المرتبة الثانية. وهذه نتيجة لاختيارات المستويات المدروسة لـ X_1 وX_2 وسيكون لنا تعليق إضافي على هذا بعد قليل.

ويمكن تحويل دالة الانحدار التوفيقية (9.18)، عائدين إلى المتغيرات الأصلية، باستخدام (9.16b) ونجد

$$\hat{Y} = 160.58 - 139.58X_1 + 7.55X_2 \qquad (9.18a)$$

ويتضمن الشكل (٩-٨) رسما ثلاثي الأبعاد من إنتاج الحاسوب لمستوى الاستجابة التوفيقي وقد استخدم الباحث سطح الاستجابة التوفيقي هذا لتحرّي تأثيرات معدل الشحن ودرجة الحرارة على حياة هذا النوع الجديد من خلايا الطاقة.

شكل (٩-٨) رسم ثلاثي الأبعاد من إنتاج الحاسوب لمستوى الاستجابة التوفيقي (9.18a) مثال خلايا الطاقة



(٩-٤) طرائقية سطح الاستجابة

إن استخدام دوال استجابة كثيرات الحدود كتقريب لسطوح استجابة معقدة هو أمر شائع في العديد من الحالات التجريبية. وقد أعطي لقب طرائقية سطح الاستجابة للطرائقية الإحصائية التي تهتم بتصميم دراسات لتقدير سطوح الاستجابة، وبالتقدير

الفعلي لسطوح الاستجابة وتفسير النتائج.

وتُستخدم طرائقية سطح الاستجابة لغرضين رئيسيين: (١) لتقديم وصف لنمط الاستجابة في منطقة المشاهدات المدروسة X، و(٢) للمساعدة في إيجاد المنطقة التي تكون الاستجابة فيها استجابة مثلى (أي حيث يكون سطح الاستجابة أعظميا أو أصغريا).

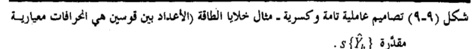
وقد ناقشنا آنفا طرق تقويم صلاحية سطح استجابة توفيقي، كما عرضنا اختبارات لتقرير ما إذا كانت تأثيرات التفاعل وتأثيرات الانحناء مطلوبة في النموذج أم لا.

وسنناقش الآن باختصار ناحيتين جديدتين لطرائقية سطح الاستجابة : (١) تصميم دراسات سطح الاستجابة، و(٢) البحث عن شروط الاستجابة المثلى.

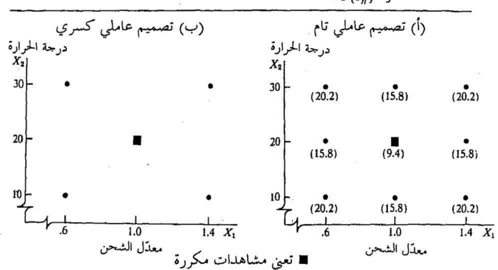
تصميم دراسات سطح استجابة

لقد تم تطوير تشكيلة كبيرة من التصاميم التجريبية لتقدير سطوح استجابة بطريقة كفؤة. ونقدم في الشكل (٩-٩)أ التصميم الخاص بمثال خلايا الطاقة.
ونلاحظ أن كلا من مستويات معدل الشحن قد اعتبر مع كل مستوى لدرجة الحرارة. ويدعى تصميم دراسة كهذه تصميما عامليا تاما. وقد كُرِّرت النقطة المركزية (1.0,20) تكرارات إضافية كي تُقدم قياسا للخطأ البحت، وللمساعدة في تقويم صلاحية نموذج توفيقي.

وعندما يكون تكرار التجربة مكلّفا ، يمكن استخدام تصميم عاملي كسري. وهنا لا ندرس كل مستوى لمتغير مستقل مع كل مستوى لكل متغير مستقل آخر. ويبين الشكل (٩-٩)ب تصميما عامليا كسريا يمكن استخدامه لمثال خلايا الطاقة. وفيه لم يُدرس تركيبات مختارة من (X_1 ، X_2) ويقدم التصميم في الشكل (٩-٩)ب معلومات عن التأثيرات الخطّية لـ X_1 وX_2 بالإضافة إلى بعض المعلومات عن تأثيرات التفاعل والانحناء. ويمكن أيضا اعتبار التصميم العاملي الكسري في الشكل (٩-٩)ب على أنه تصميم عاملي تام بمستويين لكل عامل، مع تكرارات مضافة للنقطة المركزية في التصميم.



شكل (٩-٩) تصاميم عاملية تامة وكسرية - مثال خلايا الطاقة (الأعداد بين قوسين هي انحرافات معيارية مقدّرة $s\{\hat{Y}_h\}$.



والتصميمان كلاهما في الشكل (٩-٩) قابل للتدوير. ولمثل هذه التصاميم خاصية أن الانحراف المعياري المقدَّر للقيمة التوفيقية $s\{\hat{Y}_h\}$، يبقى نفسه من أجل أي مسافة معينة من مركز التصميم، وذلك بصرف النظر عن الاتجاه. وفي الشكل (٩-٩)أ نجد الانحرافات المعيارية المقدَّرة للقيم التوفيقية في النقاط التجريبية معطاة بين قوسين، ونلاحظ أن $s\{\hat{Y}_h\}=15.8$ في كل من النقاط (1.0, 10)،(0.60, 20)، (1.4, 20)، (1.0, 30) باعتبارها تقع جميعها على مسافات متساوية عن نقطة المركز. وبصورة مماثلة فإن لجميع نقاط الزوايا انحراف معياري $s\{\hat{Y}_h\}=20.2$ لأنها جميعها متساوية البعد عن نقطة المركز وكالمعتاد ، يزداد $s\{\hat{Y}_h\}$ كلما ابتعدت النقطة التجريبية عن المركز.

وخاصية تساوي الدقة عند مسافة معطاة عن المركز في التصاميم القابلة للتدوير هي خاصية مرغوبة لأننا ، في العادة، لانعرف سلفا أي اتجاه من نقطة المركز هو الاتجاه الأهم، ويطمئننا التصميم القابل للتدوير أن دقة القيم التوفيقية لاتتأثر بالاتجاه، وإنما تتأثر بالبعد عن نقطة المركز فقط.

البحث عن شروط استجابة مثلى ـ متغير مستقل واحد

كثيرا ما نقوم بتوفيق سطوح استجابة لأغراض ايجاد شروط استجابة مثلى. وفي مثال خلايا الطاقة، على سبيل المثال، قد ترغب الإدارة بمعرفة المركّب من معدل الشحن ودرجة الحرارة الذي يجعل العمر المتوقع لخلايا الطاقة أعظم مايمكن وبصورة عامة، نحتاج إلى سلسلة من التجارب لإيجاد شروط الاستجابة المثلى، كما سنوضح الآن – أولا في حالة متغير مستقل واحد ثم في حالة متغيرين مستقلين.

وعندما ينطوي النموذج على متغير مستقل واحد فقط وتكون دالة الاستجابة تربيعية:

$$\hat{Y}=b_0+b_1x+b_{11}x^2 \tag{9.19}$$

حيث: $x=X-\overline{X}$، والقيمة العظمى (الصغرى) تقع عند المستوى x_m:

$$x_m=-\frac{b_1}{2b_{11}} \tag{9.20}$$

وبدلالة المتغير الأصلي X، تقع القيمة العظمى (الصغرى) عند المستوى X_m:

$$X_m=\overline{X}-\frac{b_1}{2b_{11}} \tag{9.20a}$$

ومتوسط الاستجابة المقدَّر عند X_m هو:

$$\hat{Y}_m=b_0-\frac{b_1^2}{4b_{11}} \tag{9.21}$$

و $\hat{Y}_m$ عظمى إذا كان b_{11} سالبا وصغرى إذا كان b_{11} موجبا.

مثال. في المثال السابق لمبيعات كافيتريا من القهوة، كان منحني الانحدار التوفيقي:

$$\hat{Y}=705.47+54.89x-4.25x^2$$

وإذا كانت دالة الانحدار التربيعية مناسبة من أجل قيم لـ x أكبر من القيم المعطاة في الدراسة، فيمكننا تقدير أن متوسط مبيعات القهوة الأعظمي يقع عند:

$$X_m=3-\frac{54.89}{2(-4.25)}=9$$

ويكون متوسط الاستجابة المقدَّر عندئذ:

$$\hat{Y}_m = 705.47 + 54.89(9) - 4.25(9)^2 = 855$$

وبما أن أكبر عدد من آلات الخدمة الذاتية في الدراسة كان $X = 6$ فقد يكون من المرغوب جدا توسيع الدراسة وتقصي مبيعات القهوة بعدد أكبر من الآلات يبلغ، مثلا 12. ويمكن عندئذ استخدام هذه المعلومات لاختبار صلاحية دالة استجابة تربيعية في المدى الأوسع للمتغير المستقل والتأكد من موقع شرط الاستجابة المثلى. وقد يكون كافيا لهذه الأغراض توسيع الدراسة بحيث تشمل $X = 8, 10, 12$ من الآلات.

تعليقات

١ـ لاستنباط (9.20) نشتق $\hat{Y}$ في (9.19) بالنسبة لـ x ونضع المشتق مساويا للصفر:

$$\frac{d\hat{Y}}{dx} = \frac{d}{dx}(b_0 + b_1 x + b_{11} x^2) = b \ \ + 2b_{11} x = 0$$

وبذلك نحصل على:

$$x_m = -\frac{b_1}{2b_{11}}$$

وبتعويض هذه القيمة في دالة الاستجابة التوفيقية (9.19)، نجد:

$$\hat{Y}_m = b_0 + b_1\left(\frac{-b_1}{2b_{11}}\right) + b_{11}\left(\frac{-b_1}{2b_{11}}\right)^2$$

$$= b_0 - \frac{b_1^2}{4b_{11}}$$

٢ـ من أجل عينات كبيرة يكون التباين التقريبي المقدَّر لـ X_m في (9.20) :

$$s^2\{X_m\} = \frac{b_1^2}{4b_{11}^2}\left[\frac{s^2\{b_1\}}{b_1^2} + \frac{s^2\{b_{11}\}}{b_{11}^2} - \frac{2s\{b_1, b_{11}\}}{b_1 b_{11}}\right] \tag{9.22}$$

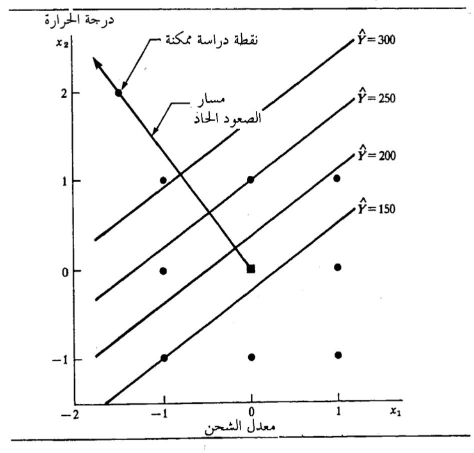
ويمكن استخدام هذا التباين المقدَّر التقريبي لوضع فترة ثقة للمستوى الحقيقي X التي تكون الاستجابة عنده عظمى (صغرى). ويمكن أيضا الحصول على فترات ثقة تقريبية لـ $E\{Y_m\}$ وقد نوقشت هذه في المرجع [9.1].

البحث عن شروط استجابة مثلى - متغيران مستقلان

عندما لايعلم الباحث سلفا المنطقة القريبة من موقع الاستجابة المثلى لمتغيرين مستقلين، فإن الاستراتيج المستخدم عادة يكون تتابعيا. ففي البداية ننفذ تصميما تجريبيا بسيطا في حيز ضيق من فضاء المتغيرين (X_1, X_2) وعلى أساس من هذه التجربة الابتدائية نحصل على معلومات عن الاتجاه الذي تقع فيه الاستجابة المثلى ويدعى هذا الاتجاه مسار الصعود (الهبوط) الحاد. ثم نقوم ببعض الأشواط التجريبية الإضافية للتعرف بصورة أوضح على موقع الاستجابة المثلى، ثم نقوم بتنفيذ تصميم تجريبي إضافي لتحديد المنطقة المثلى بدقة أكبر. وتتكرر عملية البحث هذه حتى تتحدد المنطقة المثلى بدقة كافية.

وسنستخدم مثال خلايا الطاقة لإيضاح هذه الأفكار الأساسية.

شكل (٩-١٠) خطوط تساوي الاستجابة وخط الصعود الحاد لدالة الاستجابة التوفيقية (9.18) مثال خلايا الطاقة.



مثال. يتضمن الشكل (٩-١٠) نقاط التصميم في الدراسة الابتدائية لتأثيرات معدّل الشحن ودرجة الحرارة على عمر خلايا الطاقة. ويتضمن الشكل أيضا بعض خطوط التساوي الخاصة بمستوى الاستجابة التوفيقي (9.18) ، والذي تم تحديده كتوفيق مناسب ضمن المدى المدروس للمتغيرات X.

وللبحث عن الشروط التي تجعل العمر المتوقع لخلايا الطاقة أكبر ما يمكن، نحتاج إلى دراسة الاستجابات على طول مسار الصعود الحاد. وهذا المسار متعامد مع خطوط التساوي. وباستخدام نموذج الانحدار التوفيقي (9.18) بوحدات مرمزة، نجد أنه من أجل كل $b_1 = -55.83$ من الوحدات في الاتجاه السالب على طول المحور x_1، يزداد المسار بـ $b_2 = -75.50$ من الوحدات على طول المحور x_2 وهكذا فإن ميله يساوي $-1.35 = (-55.83) / 75.50 = b_2 / b_1$ من وحدات x_2 لكل وحدة زيادة في x_1. ويبين الشكل (٩-١٠) مسار الصعود الحاد بميل يساوي 1.35 -، بدءا من نقطة مركز التصميم $(x_1=0, x_2=0)$ وإحدى نقاط الدراسة المريحة التالية يمكن أن تكون عند $x_1 = -1.5$ والقيمة المقابلة لـ x_2 على مسار الصعود الحاد هي $x_2 = -1.5(-1.35) = 2.025$. ونقطة الدراسة الممكنة هذه محددة في الشكل (٩-١٠). ويمكن الحصول على القيم الفعلية لمعدل الشحن ودرجة الحرارة عند نقطة الدراسة هذه باستخدام (9.16b) إنها $X_1 = 0.4$ و$X_2 = 40.25$.

وبعد الاستطلاع على طول خط الصعود الحاد، ينبغي تنفيذ تصميم تجريبي آخر في المنطقة التي حددتها نقاط الدراسة الإضافية كمنطقة أقرب إلى الشروط المثلى. ومن المتوقع جدا أن نحتاج عندئذ إلى نموذج مرتبة ثانية لوصف دالة الاستجابة في تلك المنطقة وصفا مناسبا. وستتكرر هذه العملية بالقدر الضروري لتحديد منطقة شروط الاستجابة المثلى تحديدا دقيقا.

ملاحظة

نناقش طرائقية سطح الاستجابة في كتب دراسية متخصصة مثل المرجع [9.2].

(٩-٥) بعض التعليقات الإضافية حول انحدار كثيرات الحدود

١- لا يخلو استخدام نماذج كثيرات الحدود في X من المآخذ. فتكلفة نماذج كهذه من درجات الحرية يمكن أن تكون أكبر مما هي في نماذج غير خطّية بديلة أو في نماذج خطّية مع تحويل المتغيرات. والمأخذ الكبير الآخر هو أنه لايمكن تجنب الخطّية المتعددة. وفي الحقيقة، يمكن أن تكون درجة الخطّية المتعددة في أعمدة المصفوفة X مرتفعة تماما إذا اقتصرت مستويات X على مدى ضيق، وذلك، على وجه الخصوص، في كثيرات الحدود من درجة مرتفعة. ولهذا السبب فقد صيغت جميع نماذج انحدار كثيرات الحدود في هذا الفصل بدلالة الانحرافات $x_i = X_i - \overline{X}$.

٢- والبديل لاستخدام متغيرات في انحدار كثيرات الحدود معبرا عنها بدلالة الانحرافات عن المتوسط، هو استخدام كثيرات الحدود المتعامدة. وكثيرات الحدود المتعامدة غير مرتبطة. وتستخدم بعض حزم الحاسب كثيرات الحدود المتعامدة في روتين انحدار كثيرات الحدود، وتقدم النتائج التوفيقية النهائية بدلالة كل من كثيرات الحدود المتعامدة وكثيرات الحدود الأصلية وتناقش كثيرات الحدود المتعامدة في كتب دراسية متخصصة مثل المرجع [9.3].

٣- يتم أحيانا توفيق دالة استجابة تربيعية لغرض إثبات خطّية دالة الاستجابة، وذلك عندما لا تتوافر مشاهدات مكررة لاختبار خطّية دالة الاستجابة بصورة مباشرة. ولكن القيام بتوفيق نموذج تربيعي:

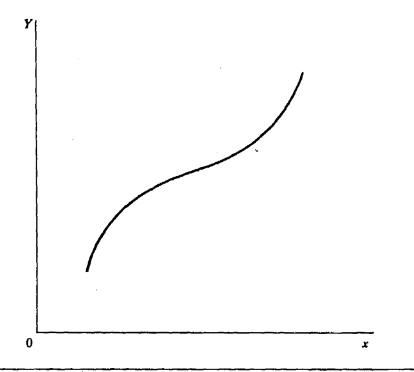
$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \varepsilon_i \qquad (9.23)$$

واختبار ما إذا كان $\beta_{11} = 0$ لايثبت بالضرورة أن دالة الاستجابة الخطّية مناسبة ويقدم الشكل (٩-١١) مثالا. إذا حصلنا على بيانات عينة من أجل دالة الاستجابة في الشكل (٩-١١) وقمنا بتوفيق النموذج (9.23)، ثم اختبرنا β_{11}، فمن المحتمل أن يقود ذلك إلى استنتاج أن $\beta_{11} = 0$ مع أنه من الواضح أن دالة الاستجابة الخطّية غير مناسبة.

٤ـ عند استخدام نموذج انحدار كثيرة حدود بمتغير مستقل واحد، نقوم عادة بتوفيق كثيرة حدود من أعلى درجة نتوقع أنها مناسبة، ويتم تفكيك SSR إلى مركبات مجاميع مربعات إضافية كما يلي:

$$SSR(x)$$
$$SSR(x^2 \mid x)$$
$$SSR(x^3 \mid x, x^2)$$
$$etc.$$

شكل (٩-١١) مثال عن دالة استجابة منحنية.



والسبب في مثل هذا الأسلوب هو أن جلّ اهتمامنا ينصبّ، بصورة عامة، على ما إذا كان يمكننا إسقاط حدود من مراتب عليا من النموذج. وهكذا فإننا عندما نقوم بتوفيق نموذج تكعيبي متوقعين أن يكون النموذج من المرتبة الثالثة كافيا، ونرغب في اختبار ما إذا كان

$\beta_{111} = 0$ أم لا، فإن مجموع المربعات الإضافي المناسب هو $SSR(x^3|x,x^2)$. وإذا رغبنا، بدلا من ذلك، في اختبار ما إذا كـان الحـد الخطّي مناسبا، أي $\beta_{11} = \beta_{111} = 0$ فـإن مجمـوع المربعات الإضافي المناسب هو$SSR(x^2,x^3|x) = SSR(x^2\ |x) + SSR(x^3\ |\ x\ ,x^2)$. وسـوف لانقوم عادة، بتوفيق نموذج من المرتبة الثالثة ونختبر أولا ما إذا كـان معـامل مـن مرتبـة أدنى صفرا أم لا، فنختبر، مثلا، ما إذا كان $\beta_{11} = 0$ أم لا. والسبب في ذلك هـو أننـا نرغب، عادة في استخدام نموذج انحدار بسيط قدر الإمكان، مما يعـني في حالـة انحـدار كثيرة حدود، نموذجا من مرتبة أدنى.

مراجع ورد ذكرها.

[9.1]Williams, E. J. *Regression Analysis*. New York : John Wiley & Sons, 1959.

[9.2]Box, G. E. P. and Draper, N. R. *Empiriecal Model-Building and Response Surfaces*. New York : John Wiley & Sons, 1987.

[9.3] Draper, N. R. and Smith, H. *Applied Regresion Analysis*, 2nd ed. New York: John Wiley & Sons, 1981.

مسائل

(٩-١) عرض متحدث مايلي :" عند تطوير نماذج انحدار كثيرة حدود من مرتبة ثالثة أو من مرتبة أعلى، في العلوم الاجتماعية والتطبيقات الإداريـة، تـأخذ الاسـتقراءات حـول المعالم β، عادة، شكل اختبارات مباشرة. وهناك اهتمام بسيط نسبيا في تقديـر المعالم β بغية تثمين تأثيرات حدود كثيرة الحدود كل بمفردها". لمـاذا يمكـن أن يكون الأمر كذلك؟

(٩-٢) ارسم عدة منحنيات تساوي لسطح الاستجابة التربيعي

$$E\{Y\} = 140 + 4x_1^2 - 2x_2^2 + 5x_1x_2$$

(٩-٣) استخدمت محللة استثمار مبتدئة نموذج انحـدار كثـيرة حـدود مـن مرتبـة عاليـة نسبيا في ندوة بحثية تتعلق بالسندات البلدية. وقد حصلت على $R^2 = 0.991$ في انحدار دخل الفائدة الصافي للسند (Y) على الرقم القياسي للتنـوع الصنـاعي في منطقة البلدية (X)، وذلك من أجل سبعة إصدارات للسندات. وقالت زميلتها، غير معجبة بالنتيجـة: "تعـاني نتـائجك مـن مبالغـة في التوفيـق والمنحـنى الـذي

حصلت عليه يخضع لتأثيرات عشوائية في البيانات".

أ - علّق على الانتقاد

ب - هل يمكن لـ R_a^2 المعرفة في (7.37) أن يكون مناسبا هنا أكثر من R^2 كمقياس وصفي؟

(٩-٤) أدخل طالب المتغيرات X في صيغة X وX^2 وذلك في عمل صفّي عن كيفية توفيق نموذج كثيرة حدود من المرتبة الثانية بمتغير مستقل واحد. وقد أقلقه أن لايأخذ برنامج الحاسوب X^2 في الاعتبار بل يحدر Y على X فقط. وقد تضمن مُخرج الحاسب الرسالة التالية:

مربع X هو متغير زائد عن الحاجة. **X'X** شاذ تقريبا عند شمول مربع X.

أوضح الموقف.

ماذا كان ينبغي على الطالب عمله؟

(٩-٥) **دراسة استهلاك سيارة للوقود.** دُرست فعالية (جير) تجريبي جديد في تخفيض استهلاك الجازولين في 12 محاولة استخدمت فيها عربة نقل خفيفة مجهزة بهذا الجير. ويرمز X_i في البيان التالي للسرعة الثابتة (بالميل في الساعة) لعربة الاختبار في المحاولة i، ويرمز لعدد الأميال المقطوعة لكل جالون.

12	11	10	9	8	7	6	5	4	3	2	1	i
60	60	55	55	50	50	45	45	40	40	35	35	X_i
30	27	37	34	39	41	38	37	31	28	20	22	Y_i

ويُتوقع أن يكون نموذج الانحدار (9.1a) من المرتبة الثانية بحدود خطأ طبيعية مستقلة، نموذجا مناسبا.

أ - قم بتوفيق نموذج الانحدار (9.1a)، ارسم دالة الانحدار التوفيقية والبيانات. هل تبدو دالة الانحدار التربيعية توفيقا جيدا هنا ؟ أوجد R^2.

ب - اختبر ما إذا كانت توجد علاقة انحدار أم لا. اضبط مخاطرة الخطأ من النوع الأول عند $\alpha = 0.05$ اعرض البدائل ، قاعدة القرار ، والنتيجة.

جـ - قدّر متوسط الأميال للجالون الواحد في أشواط للاختبار تكون السرعة

فيها 48 ميلا في الساعة. استخدم 95 بالمائة فترة ثقة. فسّر فترتك.

د ـ تنبأ بعدد الأميال في الجالون في الاختبار القادم بسرعة 48 ميلا في الساعة، استخدم 95 بالمائة فترة تنبؤ. فسّر نتيجتك.

هـ ـ اختبر ما إذا كان يمكن إسقاط الحد التربيعي من نموذج الانحدار، استخدم α =0.05. اعرض البدائل ، قاعدة القرار، والنتيجة

و ـ عبر عن دالة الانحدار التوفيقية في الجزء (آ) بدلالة المتغير الأصلي X.

ز ـ احسب معامل الارتباط البسيط بين X و X^2 وبين x و x^2 هل استخدام متغير الانحراف مفيد هنا؟ اشرح.

(٩-٦) بالإشارة إلى مسألة **استهلاك الوقود** (٩-٥).

أ ـ أوجد الرواسب وارسمها في مقابل $\hat{Y}$ وفي مقابل x في رسمين منفصلين. قم أيضا بإعداد رسم احتمال طبيعي. فسّر رسومك.

ب ـ اختبر رسميا نقص التوفيق بالنسبة لدالة الانحدار التربيعية، استخدم α =0.05. اعرض البدائل، قاعدة القرار، والنتيجة. ماهي الافتراضات الضمنية التي افترضتها في هذا الاختبار ؟

جـ ـ قم بتوفيق نموذج المرتبة الثالثة (9.3) واختبر ما إذا كان $\beta_{111} = 0$ أم لا، استخدم α =05. اعرض البدائل، قاعدة القرار و النتيجة. هل تنسجم نتيجتك مع ماوجدته في (ب)؟

(٩-٧) إنتاجية عمل على أساس القطعة. درس محلل في شركة متعددة الجنسيات للألكترونيات العوامل المؤثرة في إنتاجية عمل على أساس القطعة حيث يستند الأجر على عدد القطع المنتجة. اختبر مستخدمان من أعمار مختلفة وتمّ الحصول على انتاجيتهما في العام الماضي (X هو عمر المستخدم بالسنوات، وY إنتاجية المستخدم، كل منهما مرمّز):

i	1	2	3	4	5	6	7	8	9
X_i	20	20	25	25	30	30	35	35	40
Y_i	97	93	99	105	109	106	109	111	100

18	17	16	15	14	13	12	11	10	i
60	60	55	55	50	50	45	45	40	X_i
110	112	109	105	103	105	101	97	105	Y_i

أدرك المحلل أن العلاقة بين العمر والإنتاجية هي علاقة معقدة، وأحد الأسباب هو أن أهداف الكسب (و لم يستطع قياسها) تتغير بشكل معقد مع العمر. واعتقد، على أي حال، أنه ولأغراض تقدير متوسط الاستجابات، يمكن تقريب دالة الاستجابة تقريبا مناسبا بكثيرة حدود من المرتبة الثالثة، وأن حدود الخطأ مستقلة وتتوزع على وجه التقريب طبيعيا.

أ ـ قم بتوفيق نموذج الانحدار (9.3) ارسم دالة الانحدار التوفيقية والبيانات هل تبدو دالة الانحدار التكعيبية توفيقا جيدا هنا؟ أوجد R^2.

ب ـ اختبر ما إذا كانت توجد علاقة انحدار أم لا. استخدم مستوى معنوية 0.01 اعرض البدائل، قاعدة القرار و النتيجة. ماهي القيمة -P للاختبار؟

جـ ـ أوجد تقديرات فترة متزامنة لإنتاجية مستخدمين أعمارهم 53 ، 58 و62، على الترتيب استخدم طريقة التقدير المتزامنة الأكثر فعالية و 99 بالمائة معامل ثقة عائلي، فسّر فتراتك.

د ـ تنبأ بإنتاجية مستخدم عمره 53، مستخدما 99 بالمائة فترة تنبؤ. فسّر فترتك.

هـ ـ عبر عن دالة الانحدار التوفيقية التي حصلت عليها في الجزء (أ) بدلالة المتغير الأصلي X.

و ـ احسب معامل الارتباط البسيط بين X و X^3 وبين x و x^3. هل استخدام متغير الانحراف يفيد هنا؟ اشرح.

(٩-٨) بالإشارة إلى المسألة (٩-٧) **إنتاجية عمل على أساس القطعة.**

أ ـ اختبر ما إذا كان يمكن إسقاط كل من الحدين التربيعي والتكعيبي من نموذج الانحدار؛ استخدم $\alpha = 0.01$. اعرض البدائل ، قاعدة القرار و النتيجة.

ب ـ اختبر ما إذا كان يمكن إسقاط الحد التكعيبي بمفرده من نموذج الانحدار؛ استخدم $\alpha = 0.01$. اعرض البدائل، قاعدة القرار والنتيجة.

(٩-٩) بالإشارة إلى المسألة (٩-٧) **إنتاجية عمل على أساس القطعة.**

أ - أوجد الرواسب وارسمها في مقابل القيم التوفيقية وفي مقابل x ، في شكلين منفصلين. قم أيضا بإعداد رسم احتمال طبيعي. ماذا تبيّن رسومك؟

ب - اختبر رسميا نقص التوفيق. اضبط مخاطرة الخطأ من النوع الأول عند 0.01 اعرض البدائل، قاعدة القرار و النتيجة. ماهي افتراضاتك الضمنية في هذا الاختبار؟

(٩-١٠) **التنبؤ بالمبيعات.** أدخلت شركة ويتون (*Wheaton*) مُنتجا جديدا عام ١٩٨٠. وفيما يلي المبيعات السنوية لهذا المُنتج (Y بآلاف الوحدات)؛ الفترة الزمنية X مرمزة، حيث $X = 1$ من أجل ١٩٨٠.

9	8	7	6	5	4	3	2	1	i
9	8	7	6	5	4	3	2	1	X_i
6.44	5.99	5.56	5.12	4.73	4.41	4.05	3.78	3.49	Y_i

افترض أن نموذج انحدار كثيرة الحدود من المرتبة الثانية (9.1a) مع حدود خطأ طبيعية هو نموذج مناسب.

أ - قم بتوفيق نموذج الانحدار (9.1a). ارسم دالة الانحدار التوفيقية والبيانات. هل تبدو دالة الانحدار التربيعية توفيقا جيدا هنا ؟ ما هو R^2؟ هل تعتقد أن دالة الانحدار التربيعية مناسبة للقيام بإسقاطات حول العام 2000؟ ناقش.

ب - أوجد فترتي ثقة بونفرّوني متزامنين لـ β_1 و β_{11} بمعامل ثقة عائلي 90 بالمائة. فسّر فتراتك.

جـ - تنبأ بمبيعات المنتج عام 1990مستخدما 90 بالمائة فترة ثقة. فسّر فترتك.

د - عبّر عن دالة الانحدار التوفيقية التي حصلت عليها في الجزء (ا) بوحدات X الأصلية.

هـ - احسب معامل الارتباط البسيط بين X و X^2 وبين x و x^2، هل استخدام متغير الانحراف مفيد هنا؟ اشرح.

(٩-١١) بالإشارة إلى مسألة **التنبؤ بالمبيعات** (٩-١٠).

أ - اختبر ما إذا كان يمكن إسقاط الحد التربيعي من نموذج الانحدار. اضبط

مخاطرة الخطأ من النوع الأول عند 0.10. اعرض البدائل، قاعدة القرار، والنتيجة. ماهي القيمة -P للاختبار ؟

ب ـ أوجد الرواسب. ارسم الرواسب في مقابل القيم التوفيقية ومقابل الزمن في رسمين منفصلين. ماذا تبين رسومك؟.

(٩-١٢) **إنتاج محصول**. درس مختص في العلوم الزراعية تأثيرات الرطوبة (X_1 بالبوصة) ودرجة الحرارة (X_2 بالدرجة المئوية) على إنتاج سلالات جديدة من الطماطم Y وفيما يلي البيانات التجريبية:

9	8	7	6	5	4	3	2	1	i
8	8	8	8	6	6	6	6	6	X_i
23	22	21	20	24	23	22	21	20	X_{i2}
51.2	50.4	51.7	51.5	47.0	49.6	48.0	48.1	49.2	Y_i

18	17	16	15	14	13	12	11	10	i
12	12	12	10	10	10	10	10	8	X_i
22	21	20	24	23	22	21	20	24	X_{i2}
48.0	47.0	48.6	48.7	48.9	50.3	51.5	51.1	48.4	Y_i

25	24	23	22	21	20	19	i
14	14	14	14	14	12	12	X_i
24	23	22	21	20	24	23	X_{i2}
40.5	43.9	42.1	42.6	43.2	46.2	46.4	Y_i

ويتوقع المختص في الزراعة أن يكون نموذج انحدار كثيرات الحدود من المرتبة الثانية (9.5) بحدود خطأ طبيعية مستقلة نموذجا مناسبا هنا.

أ ـ قم بتوفيق نموذج الانحدار (9.5). ارسم المشاهدات Y مقابل القيم التوفيقية. هل تقدم دالة الاستجابة توفيقا جيدا.

ب ـ احسب R^2 ما هي المعلومات التي يقدمها هذا المقياس؟

جـ ـ اختبر ما إذا كانت توجد علاقة انحدار أم لا؛ استخدم α=0.05. اعرض البدائل، قاعدة القرار، والنتيجة ما هي القيمة -P لهذا الاختبار؟

د ـ قدّر متوسط الإنتاج عندما يكون $X_1 = 7$ و$X_2 = 22$ ؛ استخدم 95 بالمائة فترة ثقة، اعط تفسيرا لفترتك.

هـ ـ عبّر عن دالة الاستجابة التوفيقية التي حصلت عليهـا في (أ) بدلالـة المتغيرات الأصلية X.

(٩-١٣) بالإشارة إلى مسألة إنتاج **محصول** (٩-١٢).

أ ـ اختبر ما إذا كان يمكن إسقاط حد التفاعل من نموذج الانحـدار. اضبـط المخاطرة α عند 0.005. اعرض البدائل، قاعدة القرار، والنتيجة.

ب ـ مفترضا أن حد التفاعل قد أسقط من نموذج الانحدار، اختبر ما إذا كان يمكن إسقاط حد التأثير التربيعي من النموذج أم لا؛ اضبط المخاطرة α عند 0.005 اعرض البدائل، قـاعدة القـرار، والنتيجـة. مـاهي المخـاطرة المركبة α لكلى الاختبارين هنا وفي الجزء (ا) ؟

جـ ـ قم بتوفيق نموذج انحدار كثيرة حدود من المرتبـة الثانيـة حاذفـا حـد التفـاعل وحـد التأثـير الـتربيعي للدرجـة، أوجـد الرواسـب وارسمهـا مقـــابل القيـــم التوفيقية، ومقابل X_1 ومقابل X_2 في رسوم منفصلة ماذا تبين رسومك؟

(٩-١٤) **لعبة حاسوبية.** في لعبة حاسوب تسويقية اتصل بك طلاّب يمثلون شـركة A، يطلبون المساعدة في تحليل العلاقة بين النفقات التشجيعية (X) والطلب على إنتاج الشركة (Y) ضمن منطقة الشركة. ويعتقدون أن هذه العلاقـة تتصـف بالميزات التالية: (١) يتأثر الطلب في منطقة الشركة بصورة رئيسة بالنفقـات التشجيعية، (٢) العلاقة هي إما تربيعية أو خطّيـة ضمـن مـدى مسـتويات X ذات الأهمية للشركة. وقد قدم الفريق من الطلاّب البيانـات المبينـة أدناه لـ ١٤ فترة زمنية غطتها اللعبة حتى الآن (X بآلاف الـدولارات، Y بـآلاف الوحـدات) وأفادوا بأن هذه البيانات تمتد فوق جميع مستويات X ذات الأهمية.

7	6	5	4	3	2	1	i
20	15	18	10	25	15	17	X_i
55.55	55.97	56.23	52.54	55.27	54.50	56.15	Y_i

14	13	12	11	10	9	8	i
16	25	23	20	13	17	25	X_i
55.06	54.96	55.65	55.78	54.28	55.14	54.32	Y_i

افترض أن نموذج الانحدار من المرتبة الثانية (9.1a) بحدود خطأ طبيعية مستقلة هو نموذج ينطبق على هذه الحالة.

أ - قم بتوفيق هذا النموذج واختبر ما إذا كانت علاقة الانحدار موجودة أم لا. استخدم مستوى معنوية 0.01. اعرض البدائل، قاعدة القرار، والنتيجة.

ب - اختبر ما إذا كان يمكن إسقاط الحد التربيعي من النموذج أم لا. استخدم مستوى معنوية 0.01 اعرض البدائل ، قاعدة القرار ، والنتيجة.

جـ - أوجد الرواسب وارسمها مقابل $\hat{Y}$ ومقابل x ، في رسمين منفصلين. قم أيضا برسم احتمال طبيعي. ماذا تبين رسومك؟

د - قم باختبار رسمي لنقص التوفيق مستخدما مستوى معنوية 0.01. اعرض البدائل، قاعدة القرار، والنتيجة. هل تتضمن نتيجتك عدم إمكانية إدخال مزيد من التحسينات على النموذج؟ ناقش.

(٩-١٥) بالإشارة إلى مسألة **لعبة حاسوبية** (٩-١٤) يتدخل في المناقشة شخص على معرفة جيدة بلعبة التسويق الحاسوبية هذه، ويعرض بأنه في نظام المعادلات التي بنيت عليها هذه اللعبة، تقوم علاقة تربيعية بين النفقات التشجيعية وبين متوسط الطلب في منطقة الشركة. ويعتقد أن نسبة سعر مبيع الشركة إلى متوسط سعر المبيع المنافس هو متغير مهم آخر يتصل بالطلب المتوقع؛ إلا أنه لايتذكر ما إذا كان لنسبة سعري المبيع آثار خطّية وآثار تربيعية معا. ولايذكر ما إذا كانت نسبة السعرين تتفاعل مع النفقات التشجيعية في تأثيرها على الطلب. وكانت نسبة السعرين في الـ 14 فترة زمنية كما يلي:

i	1	2	3	4	5	6	7
النسبة	0.931	0.976	1.045	0.939	1.010	1.059	1.000

i	8	9	10	11	12	13	14
النسبة	0.950	0.995	1.011	1.008	0.947	1.000	1.017

أ - قم بتوفيق نموذج انحدار كثيرات الحدود من المرتبة الثانية (9.5) بنفقات تشجيعية (X_1) ونسبة سعر (X_2) كمتغيرين مستقلين. كم ازداد R^2

بإضافة نسبة السعر كمتغير مستقل؟

ب ـ اختبر ما إذا كان ينبغي الاحتفاظ بمتغير نسبة السعر في نموذج الانحدار. اضبط مخاطرة الخطأ من النوع الأول عند 0.05. اعرض البدائل، قاعدة القرار أو النتيجة.

جـ ـ مفترضا أنك ستحتفظ بمتغير نسبة السعر في نموذج الانحدار. اختبر ما إذا كنت تحتاج إلى حد التفاعل في النموذج؛ استخدم $\alpha = 0.01$ اعرض البدائل، قاعدة القرار و النتيجة. ماهي القيمة - P للاختبار؟

د ـ قرر الفريق تبني النموذج (9.5) بدون حدود تفاعل. قم بتوفيق هذا النموذج وأوجد الرواسب. ارسم الرواسب مقابل $\hat{Y}$ ومقابل الترتيب الزمني للمشاهدات في رسمين منفصلين ، قم أيضا بإعداد رسم احتمال طبيعي. فسّر هذه الرسوم واعرض ماتوصلت إليه.

(٩-١٦) بالإشارة إلى مسألة **استهلاك البنزين** (٩-٥).

أ ـ عند أية سرعة تكون دالة الاستجابة التربيعية المقدّرة في قيمتها العظمى؟ ماهو متوسط المسافة المقطوعة بالغالون عند هذه السرعة ؟.

ب ـ هل تبلغ دالة الاستجابة قيمتها العظمى ضمن مدى النموذج؟

(٩-١٧) بالإشارة إلى مسألة **التنبؤ بالمبيعات** (٩-١٠).

أ ـ في أي سنة تبلغ دالة الاستجابة التربيعية المقدّرة قيمتها الصغرى ؟ ماهو تقدير متوسط المبيعات في هذه السنة؟

ب ـ هل تبلغ دالة الاستجابة قيمتها الصغرى ضمن مدى النموذج؟

(٩-١٨) **إنتاج كيميائي حيوي.** أراد محلل مخبري إيجاد وضع الضغط (X_1) ودرجة الحرارة (X_2) الذي ينتج حصيلة أعظمية (Y) لعملية كيميائية حيوية. وقد استُخدم تصميم عاملي ابتدائي بأربع نقاط مركزية مضافة عند $X_1 = 70$ و$X_2 = 135$ وفيما يلي البيانات:

8	7	6	5	4	3	2	1	i
70	70	70	70	80	80	60	60	X_{i1}
135	135	135	135	140	130	140	130	X_{i2}
58.9	61.3	61.8	60.2	66.7	62.5	60.4	56.2	Y_i

وكان المحلل متأكدا إلى حد مـا أن نمـوذج الانحـدار مـن المرتبـة الأولى (7.1) سيكون مناسبا.

أ - أوجد دالة الاستجابة التوفيقية من المرتبة الأولى مسـتخدما المتغـيرات X المرمزة:

$$x_{i2} = (X_{i2} - 135) / 5 \text{ و } x_{i1} = (X_{i1} - 70) / 10$$

ب - حدّد مسار الصعود الحـاد مـن نقطـة التصميـم المركزيـة. ارسـم عـدّة خطوط تساوي ومسار الصعود الحاد.

جـ - قد تشكل $x_1 = 2$ نقطة دراسة تالية مريحة. حدد قيمـة x_2 المقابلـة علـى مسار الصعود الحاد وأوجد القيم الفعلية للضغط ودرجـة الحـرارة عنـد نقطة الدراسة هذه.

د - احسب التباينات المقدَّرة للقيم التوفيقية عند $x_1 = -1$ ، $x_2 = 1$ وعند $x_1 = 1$، $x_2 = -1$.

هل التباينات منسجمة مع تلك الخاصة بتصميم قابل للدوران. اشرح.

(٩-١٩) بالإشارة إلى مسألة **انتاج محصول** (٩-١٢). اعتـبر المشـاهدات الـ 15 الأولى وكأنها تشكّل نتائج دراسة استطلاعية لسطح الاستجابة. ويرغـب المختـص في العلوم الزراعية بتوفيـق نمـوذج الانحـدار مـن المرتبـة الأولى (7.1) لتحديـد مسار الصعود الحاد.

أ - أوجد دالة الاستجابة التوفيقية من المرتبة الأولى مستخدما المتغيرات المرمزة:

$$x_{i2} = (X_{i2} - 22) / 1 \text{ و } x_{i1} = (X_{i1} - 8) / 2$$

ب - حدد مسار الصعـود الحـاد مـن نقطـة التصميـم المركزيـة، ارسـم عـدّة خطوط تساوي ومسار الصعود الحاد.

جـ - أين يمكن أن تقع نقطة الدراسة المفيدة التالية ؟ اشرح.

د ـ احسب التباينات المقدّرة للقيم التوفيقية عند 1 - = x_1 ، 2- = x_2 وعند 1 = x_1، $x_2 = 2$. هل تنسجم التباينات مع تلك الخاصة بتصميم قابل للدوران ؟ اشرح.

تمارين

(٩-٢٠) اعتبر نموذج الانحدار من المرتبة الثانية بمتغير مستقل واحد ($9.1a$) والمجموعتين التاليتين من قيم X:

المجموعة ١ :	1.0	1.5	1.1	1.3	1.9	0.8	1.2	1.4
المجموعة ٢ :	12	1	123	17	415	71	283	38

احسب، لكل مجموعة، معامل الارتباط البسيط بين X و X^2 ثم بين x و x^2 احسب أيضا معامل الارتباط بين X و X^3 وبين x و x^3 ما هي التعميمات التي تقترحها نتائجك؟.

(٩-٢١) (تحتاج إلى حساب التفاضل) بالإشارة إلى دالة الاستجابة من المرتبة الثانية (9.2) اشرح بدقة معنى معامل التأثير الخطّي β_1 ومعامل التأثير التربيعي β_{11}.

(٩-٢٢) أ ـ استنبط عبارات لِ b_0'، b_1' و b_{11}' في (9.15).

ب ـ مستخدما النظرية (6.47) أوجد مصفوفة التباين ـ التغاير لمعاملات الانحدار المتعلقة بالمتغيرات X الأصلية وذلك بدلالة مصفوفة التباين ـ التغاير لمعاملات الانحدار المتعلقة بالمتغيرات x بعد التحويل.

(٩-٢٣) كيف يمكن تبسيط المعادلات الناظمية (9.12) إذا كانت المتغيرات X على مسافات متساوية بعضها عن بعض، مثل تمثيل السلاسل الزمنية $(X_1 = 1, X_2 = 2, \ldots, X_n = n)$؟

مشاريع

(٩-٢٤) بالإشارة إلى مجموعة البيانات **SMSA** نرغب في توفيق نموذج الانحدار من المرتبة الثانية (9.1a) لإيجاد علاقة بين عدد الأطباء الممارسين (Y) وعدد السكان الإجمالي (X).

أ ـ قم بتوفيق نموذج انحدار من المرتبة الثانية، ارسم الرواسب مقابل القيم التوفيقية. كيف يبدو نموذج المرتبة الثانية من حيث جودة توفيقه للبيانات؟

ب ـ أوجد R^2 لنموذج انحدار المرتبة الثانية. أوجد أيضا معامل الارتباط البسيط لنموذج انحدار المرتبة الأولى. هل تزيد إضافة الحد التربيعي في نموذج الانحدار معامل التحديد زيادة كبيرة ؟

جـ ـ اختبر ما إذا كان يمكن إسقاط الحد التربيعي من نموذج الانحدار؛ استخدم $\alpha = 0.05$. اعرض البدائل، قاعدة القرار، والنتيجة.

د ـ احذف المشاهدة 1 (مدينة نيويورك) من مجموعة البيانات. وقم بتوفيق نموذج انحدار المرتبة الثانية (9.1a) مستندا إلى مشاهدات الـ **SMSA** الـ 140 الباقية. أعد الاختبار في الجزء (جـ). هل أثر إلغاء المشاهدة القاصية في نتيجتك حول ما إذا كان يمكن إسقاط الحد التربيعي من النموذج؟

(٩ـ٢٥) بالإشارة إلى مجموعة بيانات **SMSA**، نريد إقامة نموذج انحدار يربط بين معدل الجرائم الخطرة (Y، العددالكلي للجرائم الخطرة مقسوما على عدد السكان الكلي) وبين الكثافة السكانية (X_1، عدد السكان الكلي مقسوما على المساحة)، والنسبة المئوية للسكان في مدن مركزية (X_3).

أ ـ قم بتوفيق نموذج انحدار المرتبة الثانية (9.5) ارسم الرواسب مقابل القيم التوفيقية. كيف يبدو نموذج المرتبة الثانية من حيث جودة توفيقه للبيانات؟ ما هي R^2؟.

ب ـ اختبر ما إذا كان يمكن إسقاط جميع حدود التفاعل والحدود التربيعية من النموذج أم لا؟ استخدم $\alpha = 0.01$. اعرض البدائل، قاعدة القرار، والنتيجة.

جـ ـ بدلا من استخدام الكثافة السكانية كمتغير مستقل، نريد استخدام عدد السكان الكلي (X_1)، ومساحة المنطقة المسكونة (X_2). كمتغيرين مستقلين منفصلين، بالإضافة إلى النسبة المئوية للسكان في مدن مركزية (X_3). وينبغي أن يتضمن نموذج الانحدار حدودا خطّية وتربيعية في

عدد السكان الكلي، وحدودا خطّية فقط في مساحة المنطقة المسكونة، وفي النسبة المئوية للسكان في مدن مركزية. (لاحدود تفاعل في هذا النموذج) قم بتوفيق نموذج الانحدار هذا واحسب R^2. هل يختلف معامل التحديد المتعدد هذا اختلافا كبيرا عن المعامل الخاص بنموذج الانحدار في الجزء (أ)؟

(٩-٢٦) بالإشارة إلى مجموعة البيانات نريد توفيق نموذج الانحدار من المرتبة الثانية (9.1a) لإيجاد علاقة بين عدد المرضات Y وبين الخدمات والتسهيلات المتوافرة X.

أ - قم بتوفيق نموذج انحدار من المرتبة الثانية. ارسم الرواسب مقابل القيم التوفيقية. هل يبدو نموذج المرتبة الثانية ملائما من حيث جودة توفيقه للبيانات؟.

ب - احسب R^2 لنموذج انحدار المرتبة الثانية. واحسب أيضا معامل الارتباط البسيط r^2 لنموذج انحدار المرتبة الأولى. هل تزيد، إضافة الحد التربيعي في نموذج الانحدار، معامل التحديد زيادة كبيرة؟

جـ - اختبر ما إذا كان يمكن إسقاط الحد التربيعي من نموذج الانحدار؛ استخدم $\alpha = 0.10$. اعرض البدائل، قاعدة القرار، والنتيجة.

(٩-٢٧) بالإشارة إلى مسألة **التنبؤ بالمبيعات** رقم (٩-١٠). بدلا من استخدام نموذج انحدار كثيرات الحدود هنا، اقترح: أن تحويل المتغيرات يمكن أن يؤدي إلى توفيق بالجودة نفسها ويكون مرغوبا أكثر من حيث إن التنبؤ يتطلب التمديد (أو الاستيفاء الخارجي).

أ - قم بتوفيق نموذج انحدار يربط بين $Y' = \sqrt{Y}$ وX. ارسم دالة الانحدار التوفيقية والبيانات بعد التحويل. كيف تبدو فعالية استخدام المتغير بعد التحويل هنا؟

ب - أوجد القيم التوفيقية وحوّلها عائدا إلى المتغير الأصلي Y احسب الرواسب في المتغير الأصلي وارسم هذه الرواسب مقابل X. فسّر رسمك.

جـ ـ قم بتربيع الرواسب التي حصلت عليها في الجزء (ب)، اجمعها لتحصل على MSE قارنه مع MSE لنموذج الانحدار التربيعي في المسألة (٩ـ١٠)أ. كيف تقارن التغير حول دالة الانحدار التوفيقية، مقاسا بـ MSE في الأسلوبين؟.

د ـ أعد الأجزاء من (أ) إلى (جـ) مستخدما التحويل $Y' = log_{10}Y$. هل تتضح أفضلية تحويل الجذر التربيعي أو التحويل اللوغاريتمي هنا؟.

المتغيرات المستقلة النوعية

استخدمنا المتغيرات الكمّيّة في نماذج الانحدار المدروسة في الفصول السابقة عن تحليل الانحدار. وتأخذ المتغيرات الكمية قيما على مقياس معرف تعريفا جيدا. ومن أمثلة ذلك الدخل، العمر، درجة الحرارة وفقدان الممتلكات المنقولة.

وعلى أي حال فكثير من المتغيرات المهمة في الأعمال، الاقتصاد، العلوم الاجتماعية والأحيائية لاتكون دائما متغيرات كمية، ولكنها دائما نوعية (وصفية) ومن أمثلة تلك المتغيرات النوعية الجنس (ذكر، أنثى)، الموقف من شراء سلعة (شراء، عدم شراء) وحالات الإعاقة (غير عاجز، عاجز جزئيا، عاجز كليا).

ويمكن استخدام المتغيرات النوعية في الانحدار المتعدد. وسنأخذ في هذا الفصل الحالة التي تكون فيها بعض المتغيرات المستقلة أو كلها متغيرات نوعية.

(١٠-١) متغير نوعي واحد مستقل

يرغب أحد الاقتصاديين في ربط السرعة Y التي تم بها تبنّي شركة تأمين لخطة تأمين مبتكرة بحجم شركة التأمين (X). وأيضا بنوع الشركة. ويُقاس المتغير التابع بعدد الشهور المنصرمة بين الوقت الذي تبنّت فيه أول شركة ذلك الابتكار والوقت الذي تبنّت فيه الشركة المعنية ذلك الابتكار. إن المتغير المستقل الأول وهو حجم الشركة متغير كمي، ويقاس بكمية الأصول الكلية للشركة، أما المتغير المستقل الثاني وهو نوع الشركة فمتغير نوعي، ويتكون من صفين: شركات مساهمة، وشركات تعاونية. ولاستخدام مثل هذا المتغير في نماذج الانحدار لابد من وضع مؤشرات كمية لصفوف ذلك المتغير النوعي.

متغيرات مؤشرة

هناك عدة طرق لتعريف صفوف متغير نوعي بطريقة كمية، وسوف نستخدم متغيرات مؤشرة تتخذ القيم 0 و1. هذه المتغيرات المؤشرة سهلة الاستعمال وتستخدم

بكثرة، ولكنها ليست، بأي حال من الأحوال، الطريقة الوحيدة لتكميم متغير نوعي.

وفي مثال ابتكارات التأمين حيث يوجد صفّان للمتغير النوعي يمكن أن نُعرِّف متغيرين مؤشرين X_2 وX_3 كما يلي:

$$X_2 \begin{cases} = 1 & \text{إذا كانت الشركة مساهمة} \\ = 0 & \text{فيما عدا ذلك} \end{cases} \tag{10.1}$$

$$X_3 \begin{cases} = 1 & \text{إذا كانت الشركة تعاونية} \\ = 0 & \text{فيما عدا ذلك} \end{cases}$$

وبفرض أن النموذج المستخدم من المرتبة الأولى نجد:

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \tag{10.2}$$

حيث: $X_{i0} \equiv 1$

وللأسف فإن هذه الطريقة البدهية لوضع متغير مؤشر لكل صف من صفوف المتغير النوعي تؤدي إلى صعوبات حسابية. ولرؤية ذلك افترض أن لدينا $n=9$ مشاهدات، الأوليتان مساهمتان حيث $X_2=1$ و$X_3=0$ والأخريتان تعاونيتان حيث $X_2=0$ و $X_3=1$.

فتكون المصفوفة **X** عندئذ كما يلي:

$$\mathbf{X} = \begin{array}{c} \begin{matrix} X_0 & X_1 & X_2 & X_3 \end{matrix} \\ \begin{bmatrix} 1 & X_{11} & 1 & 0 \\ 1 & X_{21} & 1 & 0 \\ 1 & X_{31} & 0 & 1 \\ 1 & X_{41} & 0 & 1 \end{bmatrix} \end{array}$$

لاحظ أن العمود X_0 مساوٍ لمجموع العمودين X_2 وX_3، ولذلك فهذه الأعمدة ليست مستقلة طبقا للتعريف (6.22)، مما يؤدي إلى تأثيرات خطيرة على المصفوفة **X'X**.

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ X_{11} & X_{21} & X_{31} & X_{41} \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & X_{11} & 1 & 0 \\ 1 & X_{21} & 1 & 0 \\ 1 & X_{31} & 0 & 1 \\ 1 & X_{41} & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & \sum_{i=1}^{4} X_{i1} & 2 & 2 \\ \sum_{i=1}^{4} X_{i1} & \sum_{i=1}^{4} X_{i1}^2 & \sum_{i=1}^{2} X_{i1} & \sum_{i=3}^{4} X_{i1} \\ 2 & \sum_{i=1}^{2} X_{i1} & 2 & 0 \\ 2 & \sum_{i=3}^{4} X_{i1} & 0 & 2 \end{bmatrix}$$

يتضح بسرعة أن العمود الأول للمصفوفة **X'X** يساوي مجموع العمودين الأخيرين ولذلك فالأعمدة ليست مستقلة وبالتالي ليس للمصفوفة **X'X** مقلوب، ولايمكن إيجاد تقديرات وحيدة لمعاملات الانحدار.

وللخروج من هذه الصعوبة نستبعد ببساطة أحد تلك المتغيرات المؤشرة. ففي مثالنا يمكن أن نستبعد X_3، وهذا الاستبعاد ليس للخروج من الصعوبة فقط وإنما يؤدي أيضا إلى تفسيرات بسيطة للمعالم. وبصورة عامة سنتبع إذن المبدأ التالي:

(10.3) سيُمثّل المتغير النوعي الذي له c من الصفوف بـ $c - 1$ من المتغيرات المؤشرة التي تأخذ إحدى القيمتين 0 أو 1.

ملاحظة

يُطلق على المتغيرات المؤشرة متغيرات الدُّمي، أو المتغيرات الثنائية، ويشير المصطلح الأخير إلى النظام العددي الثنائي الذي يتضمن الرقمين 0 و1 فقط.

تفسير معاملات الانحدار

بالعودة إلى مثال ابتكارات التأمين أسقطنا المتغير المؤشر X_3 من نموذج الانحدار (10.2) ليصبح النموذج:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad (10.4)$$

حيث: X_{i1} = حجم الشركة

$$X_{i2} \quad \begin{array}{ll} = 1 & \text{إذا كانت الشركة مساهمة} \\ = 0 & \text{فيما عدا ذلك} \end{array}$$

وتكون دالة الاستجابة لنموذج الانحدار هذا :

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (10.5)$$

وكي نفهم معنى معاملات الانحدار لهذا النموذج، لنعتبر أولا حالة شركة تعاونية فلمثل هذه الشركة يكون $X_2 = 0$ ولدينا النموذج:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2(0) = \beta_0 + \beta_1 X_1 \qquad \text{شركة تعاونية} \qquad (10.5a)$$

أي أن دالة الاستجابة للشركة التعاونية هو خط مستقيم بمقطوع من المحور Y يساوي β_0، وميل يساوي β_1. ودالة الاستجابة هذه موضّحة في شكل (١٠-١). وفي حالة شركة مساهمة تكون $X_2 = 1$ وتصبح دالة الاستجابة (10.5) :

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2(1) = (\beta_0 + \beta_2) + \beta_1 X_1 \qquad \text{شركة مساهمة} \qquad (10.5b)$$

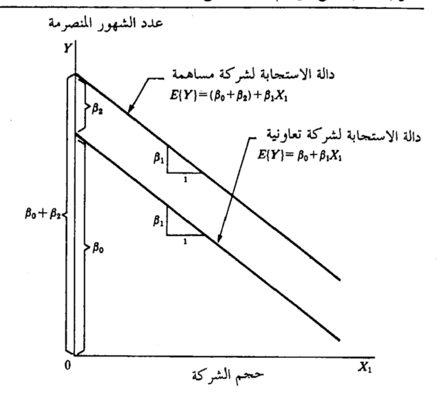
وهذا أيضا خط مستقيم بالميل نفسه β_1، ولكن بمقطوع من المحور Y هو $\beta_0 + \beta_2$. ودالة الاستجابة هذه موضّحة أيضا في الشكل (١٠-١). ويتضح الآن معنى معاملات الانحدار في دالة الاستجابة (10.5).

وبالإشارة إلى مثال خطة التأمين المبتكرة يكون $E(Y)$ متوسط الوقت المنصرم منذ تبني الابتكار للمرة الأولى، دالة خطّية في حجم المنشأة (X_1) بالميل نفسه β_1 لكل من نوعي الشركات. وتشير β_2 إلى أي مدى تكون دالة الاستجابة أعلى (أدنى) في حالة الشركات المساهمة منها في حالة الشركات التعاونية. ولذلك فإن β_2 تقيس الفرق في التأثير الناتج عن نوع الشركة. وبصورة عامة، يوضح β_2 إلى أي مدى يكون متوسط خط الاستجابة للصف المرمَّز ١ أعلى (أدنى) من خط الاستجابة للصف المرمَّز 0 وذلك لأي مستوى معطى لـ X_1.

مثال

في مثال خطّة التأمين المبتكرة، قام الاقتصادي بدراسة 10 شركات تعاونية و 10 شركات مساهمة. والبيانات موضّحة في الجدول (١٠-١). ويتضمن الجدول (١٠-٢) مصفوفتي البيانات **Y** و **X**. لاحظ أن $X_2 = 1$ لكل شركة مساهمة و$X_2 = 0$ لكل شركة تعاونية.

شكل (١٠-١) توضيح لمعنى معالم الانحدار للنموذج (10.4) بمتغير مؤشر X_2 - مثال خطة التأمين المبتكرة



وبمعرفة المصفوفتين **Y** و**X** من الجدول (١٠-٢) يصبح توفيق نموذج الانحدار (10.4) أمرا واضحا. ويقدّم الجدول (١٠-٣) النتائج الرئيسة الصادرة عن حاسوب. ودالة الاستجابة الناتجة هي:

$$\hat{Y} = 33.387407 - 0.10147X_1 + 8.05547X_2$$

ويحتوي الشكل (١٠-٢) على دالة الاستجابة التوفيقية لكل نوع من الشركات بالإضافة إلى المشاهدات الفعلية.

وكان أكثر مايثير الاهتمام هو تأثير نوع الشركة X_2 على الوقت المنصرم حتى يُؤخَذَ بالابتكار. ولذلك فهو يرغب في الحصول على 95 في المائة فترة ثقة لـ β_2 ولهذا يحتاج للقيمة $t(.975; 17) = 2.110$ ونحصل من الجدول (١٠-٣)، على حدي الثقة $8.05547 \pm 2.110\ (1.45911)$، وتكون فترة الثقة لـ β_2 هي:

$$4.98 \le \beta_2 \le 11.13$$

جدول (١٠-١) بيانات مثال خطة تأمين مبتكرة

الشركة i	عدد الأشهر المنصرمة Y_i	الشركة (بملايين الدولار الحجم) X_{i1}	نوع الشركة
1	17	151	تعاونية
2	26	92	تعاونية
3	21	175	تعاونية
4	30	31	تعاونية
5	22	104	تعاونية
6	0	277	تعاونية
7	12	210	تعاونية
8	19	120	تعاونية
9	4	290	تعاونية
10	16	238	تعاونية
11	28	164	مساهمة
12	15	272	مساهمة
13	11	295	مساهمة
14	38	68	مساهمة
15	31	85	مساهمة
16	21	224	مساهمة
17	20	166	مساهمة
18	13	305	مساهمة
19	30	124	مساهمة
20	14	246	مساهمة

جدول (١٠-٢) مصفوفات بيانات مثال خطة تأمين مبتكرة

$$\mathbf{Y} = \begin{bmatrix} 17 \\ 26 \\ 21 \\ 30 \\ 22 \\ 0 \\ 12 \\ 19 \\ 4 \\ 16 \\ 28 \\ 15 \\ 11 \\ 38 \\ 31 \\ 21 \\ 20 \\ 13 \\ 30 \\ 14 \end{bmatrix} \quad \mathbf{X} = \overset{\begin{matrix} X_0 & X_1 & X_2 \end{matrix}}{\begin{bmatrix} 1 & 151 & 0 \\ 1 & 92 & 0 \\ 1 & 175 & 0 \\ 1 & 31 & 0 \\ 1 & 104 & 0 \\ 1 & 277 & 0 \\ 1 & 210 & 0 \\ 1 & 120 & 0 \\ 1 & 290 & 0 \\ 1 & 238 & 0 \\ 1 & 164 & 1 \\ 1 & 272 & 1 \\ 1 & 295 & 1 \\ 1 & 68 & 1 \\ 1 & 85 & 1 \\ 1 & 224 & 1 \\ 1 & 166 & 1 \\ 1 & 305 & 1 \\ 1 & 124 & 1 \\ 1 & 246 & 1 \end{bmatrix}}$$

وهكذا نستنتج وبثقة 95 في المائة، ولأي حجم معطى للشركة، أن الشركات المساهمة تنحو إلى التأخر عن الشركات التعاونية في تبنّي الابتكار فترة تتراوح في المتوسط ما بين 5 أشهر و 11 شهرا.

جدول (١٠-٣) نتائج الانحدار لتوفيق النموذج (10.4) - مثال خطة تأمين مبتكرة

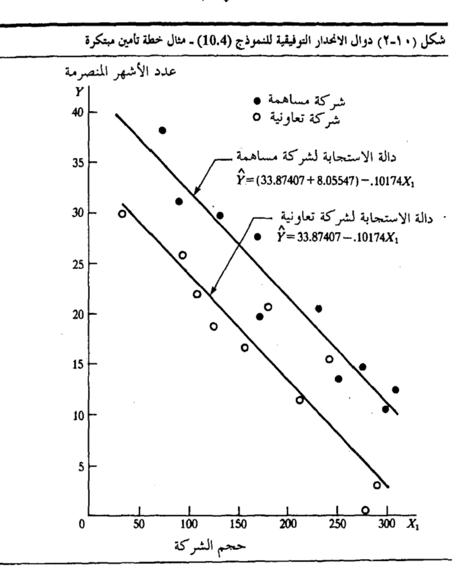
(أ) معاملات الانحدار

معامل الانحدار	معامل الانحدار المقدر	الانحراف المعياري المقدر	t^*
β_0	33.87407	1.81386	18.68
β_1	−0.10174	0.00889	−11.44
β_2	8.05547	1.45911	5.52

(ب) تحليل التباين

مصدر التغير	*SS*	*df*	*MS*
الانحدار	1,504.41	2	752.20
الخطأ	176.39	17	10.38
المجموع	1,680.80	19	

شكل (١٠-٢) دوال الانحدار التوفيقية للنموذج (10.4) - مثال خطة تأمين مبتكرة



والاختبار الرسمي التالي:

$$H_0 : \beta_2 = 0$$
$$H_a : \beta_2 \neq 0$$

بمستوى معنوية 0.05 سيؤدي إلى H_a، أي أن لنوع الشركة فيما يبدو تأثيرا ، ذلك لأن الـ 95 بالمائة فترة ثقة لـ β_2 لا تتضمن الصفر.

قام الاقتصادي بإجراء تحليل آخر وسنصف بعضا منه بعد قليل.

ملاحظة

قد يعجب القارىء لماذا لم نقم ببساطة بتوفيق خطّي انحدار منفصلين لكل من الشركات المساهمة والشركات التعاونية في مثالنا، بل قمنا بدلا من ذلك بتبنّي أسلوب توفيق انحدار واحد مع استخدام متغير مؤشّر. وهناك سببان لهذا. فبما أن النموذج يفترض تساوي الميلين كما يفترض أن التباين الثابت لحد الخطأ هو نفسه في كل من نوعي الشركات، فإن أفضل ما يمكن القيام به لتقدير الميل المشترك β_1 هو دمج النوعين من الشركات.ويمكن أيضا الحصول على استقراءات أخرى مثل تلك المتعلقة بـ β_0 و β_2 بطريقة أكثر دقة في حال استخدام نموذج انحدار واحد يحتوي على متغير مؤشر، ذلك لأن عدد درجات الحرية الموافق لمتوسط مربعات الخطأ سيكون عندئذ أكبر.

(١٠-٢) نموذج يحتوي على تأثيرات تفاعل

في مثال خطّة التأمين المبتكرة لم يبدأ الاقتصادي تحليله في الواقع بالنموذج (10.4) إذ توقّعَ توقع تأثيرات للتفاعل بين حجم المنشأة ونوعها. ومع أن أحد المتغيرات المستقلة في نموذج الانحدار متغير نوعي، فقد أُدخلت تأثيرات التفاعل في النموذج، وذلك بوضع حدود جدائية في النموذج. وفيما يلي نموذج من المرتبة الأولى مع حد تفاعل:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \qquad (10.6)$$

حيث: حجم الشركة $= X_{i1}$

$X_{i2} =$ 1 إذا كانت الشركة مساهمة

0 فيما عدا ذلك

ودالة الاستجابة لهذا النموذج هي:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \qquad (10.7)$$

معنى معاملات الانحدار

يمكن فهم معاملات الانحدار في دالة الاستجابة (10.7) على أحسن وجه باختبار طبيعة تلك الدالة لكل من نوعي الشركات. ففي الشركات التعاونية $X_2 = 0$ ولذا فإن $X_1 X_2 = 0$ وتكون دالة الاستجابة للشركات التعاونية:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2(0) + \beta_3(0) = \beta_0 + \beta_1 X_1 \qquad (10.7a)$$ شركات تعاونية

ودالة الاستجابة هذه مبينة في الشكل (١٠-٣)، لاحظ أن الجزء المقطوع من المحور Y هو β_2 وأن الميل β_1، وذلك لدالة الاستجابة للشركات التعاونية.

أما للشركات المساهمة فلدينا $X_2 = 1$، ولذا فإن $X_1X_2 = X_1$ ودالة الاستجابة للشركات المساهمة هي كالتالي:

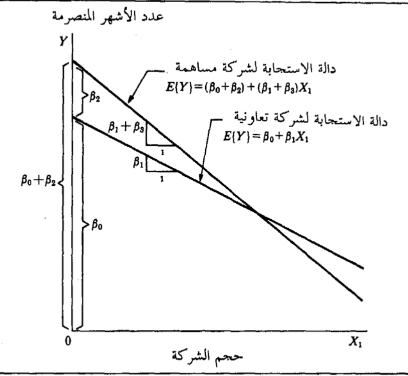
$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2(1) + \beta_3 X_1$$

أو:

$$E\{Y\} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_1 \quad \text{شركات مساهمة} \tag{10.7b}$$

ويوضح الشكل (١٠-٣) هذه الدالة أيضا. لاحظ في دالة الاستجابة للشركات المساهمة أن الجزء المقطوع من المحور Y هو $\beta_0 + \beta_2$ والميل $\beta_1 + \beta_3$.

شكل (١٠-٣) توضيح لمعنى معالم الانحدار للنموذج (10.6) بمتغير مؤشر X_2 وحد تفاعل، مثال خطة تأمين مبتكرة.

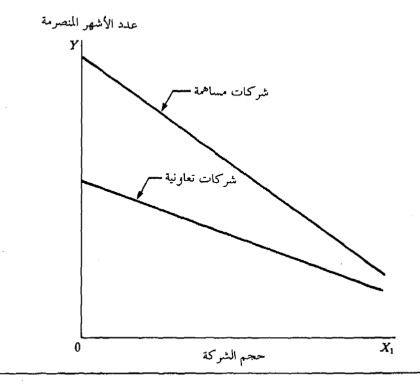


وهكذا تشير β_2 إلى أي مدى يكون الجزء المقطوع من المحور Y أكبر (أصغر) للصف المرمَّز 1 منه للصف المرمَّز 0. وبالمثل تشير β_3 إلى أي مدى يكون الميل أكبر (أصغر) للصف المرمَّز 1 منه للصف المرمَّز 0. ولأن كلا من الجزء المقطوع والميل مختلفان في نموذج الانحدار (10.6) فإنه لم يعد صحيحا أن β_2 تشير إلى أي مدى يكون

أحد خطي الاستجابة أعلى (أدنى) من الآخر لأي مستوى معطى من X_1. ويوضح الشكل (١٠-٣) أن تأثير نوع الشركة لنموذج الانحدار (10.6) يعتمد على حجم الشركة X_1 . ففي حالة الشركات الصغيرة، ووفقا للشكل (١٠-٣)، تكون الشركات التعاونية أسرع إلى تبنّي خطّة مبتكرة، أما بالنسبة للشركات الكبيرة فإن الشركات المساهمة هي الأسرع. ولهذا فإنه مع وجود تفاعل، يمكن دراسة تأثير المتغير النوعي فقط بمقارنة دوال الانحدار لكل صف من صفوف المتغير النوعي.

ويوضح الشكل (١٠-٤) سلوك أحد حالات التفاعل الممكنة لمثال خطّة تأمين مبتكرة. إذ تتجه الشركات التعاونية إلى الأخذ بالخطّة المبتكرة بصورة أسرع من الشركات المساهمة، أيا كان حجم الشركة وذلك في المجال الذي تناوله النموذج. ولكن التأثير التفاضلي للنوع أصغر بكثير في الشركات الكبيرة منه في الشركات الصغيرة.

شكل (١٠-٤) توضيح آخر للنموذج (10.6) بمتغير مؤشر X_2 وَحَدّ تفاعل - مثال خطّة تأمين مبتكرة.



مثال

بما أن الاقتصادي توقّع إمكانية وجود تأثيرات تفاعل بين حجم ونوع الشركة. فقد رغب في الواقع في توفيق النموذج (10.6):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

ويبين الجدول (١٠-٤) مصفوفة **X** لهذا النموذج. أما المصفوفة **Y** فهي نفسها كما في الجدول (١٠-٢). لاحظ أن عمود $X_1 X_2$ في المصفوفة **X** في الجدول (١٠-٤) يحتوي على 0 للشركات التعاونية و X_{i1} للشركات المساهمة.

ومع معرفة المصفوفتين **Y** و**X** يصبح توفيق النموذج أمرا روتينيا. والنتائج الأساسية الصادرة عن الحاسب مبينة في جدول (١٠-٥) ولاختبار وجود تأثيرات تفاعل:

$$H_0 : \beta_3 = 0$$
$$H_a : \beta_3 \neq 0$$

فقد استخدم الاقتصادي الإحصاءة t^* من الجدول (١٠-٥)أ:

$$t^* = \frac{b_3}{s\{b_3\}} = \frac{-0.0004171}{0.01833} = -0.02$$

وعند مستوى معنوية 0.05، نحتاج إلى $t(.975;16) = 2.120$. وبما أن $|t^*| = 0.20$ أقل من 2.12، فإننا نقبل H_0 أي $\beta_3 = 0$، ونستنتج عدم وجود تأثيرات تفاعل. وبما أن القيمة ‒P للاختبار ثنائي الجانب عالٍ جدا وهي 0.98 فقد اختار الاقتصادي نموذج الانحدار (10.4) بدون حد تفاعل. وهذا النموذج هو الذي ناقشناه في البداية.

ملاحظة

يؤدي توفيق نموذج الانحدار (10.6) إلى دوال الاستجابة نفسها كما لو وفقنا نموذجي انحدار منفصلين للشركات المساهمة والشركات التعاونية. وميزة استخدام النموذج (10.6) بمتغير مؤشّر أنه يُنتج بتشغيلة انحدار واحدة للحاسب المعادلتين التوفيقيتين لخطي الانحدار.

وميزة أخرى هي أنه يمكن بوضوح رؤية الاختبارات الخاصة بمقارنة دوال الانحدار للصفوف المختلفة للمتغير النوعي على أنها اختبارات حول معاملات الانحدار في نموذج خطّي عام. ويوضّح الشكل (١٠-٣) في مثال خطّة تأمين مبتكرة أن اختبار ما إذا كان لدالتي الانحدار الميل نفسه منطوي على اختبار:

$$H_0 : \beta_3 = 0$$
$$H_a : \beta_3 \neq 0$$

جدول (١٠-٤) المصفوفة X لتوفيق النموذج (10.6) مع حدّ تفاعل - مثال خطة تأمين مبتكرة

$$
\begin{array}{cc}
 & \begin{array}{cccc} X_0 & X_1 & X_2 & X_1X_2 \end{array} \\
\mathbf{X} = & \begin{bmatrix}
1 & 151 & 0 & 0 \\
1 & 92 & 0 & 0 \\
1 & 175 & 0 & 0 \\
1 & 31 & 0 & 0 \\
1 & 104 & 0 & 0 \\
1 & 277 & 0 & 0 \\
1 & 210 & 0 & 0 \\
1 & 120 & 0 & 0 \\
1 & 290 & 0 & 0 \\
1 & 238 & 0 & 0 \\
1 & 164 & 1 & 164 \\
1 & 272 & 1 & 272 \\
1 & 295 & 1 & 295 \\
1 & 68 & 1 & 68 \\
1 & 85 & 1 & 85 \\
1 & 224 & 1 & 224 \\
1 & 166 & 1 & 166 \\
1 & 305 & 1 & 305 \\
1 & 124 & 1 & 124 \\
1 & 246 & 1 & 246
\end{bmatrix}
\end{array}
$$

جدول (١٠-٥) نتائج انحدار لتوفيق النموذج (10.6) مع حدّ تفاعل - مثال خطة تأمين مبتكرة.

(أ) معاملات الانحدار

معامل الانحدار	معامل الانحدار المقدّر	الانحراف المعياري المقدّر	t^*
β_0	33.83837	2.44065	13.86
β_1	−0.10153	0.01305	−7.78
β_2	8.13125	3.65405	2.23
β_3	−0.0004171	0.01833	−0.02

(ب) تحليل التباين

مصدر التغير	SS	df	MS
الانحدار	1,504.42	3	501.47
الخطأ	176.38	16	11.02
المجموع	1,680.80	19	

وبالمثل فإن اختبار ما إذا كانت دالتا الانحدار متطابقتين سينطوي على اختبار:

$H_0 : \beta_2 = \beta_3 = 0$

$H_a :$ ليست كل من β_2 و β_3 مساويا للصفر

(١٠-٣) نماذج أكثر تعقيدا

نعتبر الآن باختصار، نماذج تتضمن متغيرات مستقلة نوعية وأكثر تعقيدا.

متغيرات نوعية بأكثر من صفين

إذا كان للمتغير النوعي المستقل أكثر من صفين فإن ذلك يتطلب متغيرات مؤشرة إضافية في نموذج الانحدار. لندرس انحدار اهتراء آلة معينة Y على سرعة الآلة X_1 آخذين في الاعتبار طراز الآلة (M1 , M2, M3, M4) كمتغير مستقل أيضا، وحيث إن للمتغير النوعي، طراز الآلة، أربعة صفوف فإننا نحتاج إلى متغيرات مؤشرة ـ دعنا نعرّفها كما يلي:

$$X_2 = 1 \quad \text{إذا كان طراز الآلة M1}; \quad = 0 \quad \text{فيما عدا ذلك}$$
$$X_3 = 1 \quad \text{إذا كان طراز الآلة M2}; \quad = 0 \quad \text{فيما عدا ذلك} \qquad (10.8)$$
$$X_4 = 1 \quad \text{إذا كان طراز الآلة M3}; \quad = 0 \quad \text{فيما عدا ذلك}$$

نموذج من المرتبة الأولى. نموذج الانحدار من المرتبة الأولى في هذه الحالة هو:

$$Y_i\{Y\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i \qquad (10.9)$$

والبيانات للمصفوفة **X** في هذا النموذج هي كما يلي:

طراز الآلة	$\mathbf{X_0}$	$\mathbf{X_1}$	$\mathbf{X_2}$	$\mathbf{X_3}$	$\mathbf{X_4}$
M1	1	X_{i1}	1	0	0
M2	1	X_{i1}	0	1	0
M3	1	X_{i1}	0	0	1
M4	1	X_{i1}	0	0	0

ودالة الاستجابة لنموذج الانحدار (10.9) هي:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \qquad (10.10)$$

ولكي ترى معنى معاملات الانحدار، اعتبر أولا دالة الاستجابة لآلات الطراز

M4 حيث $X_2 = 0$، $X_3 = 0$ و $X_4 = 0$:

آلات الطراز M4 $$E\{Y\} = \beta_0 + \beta_1 X_1 \qquad (10.10a)$$

ولآلات الطراز M1 يكون $X_2 = 1$، $X_3 = 0$ و $X_4 = 0$: ودالة الاستجابة هي:

آلات الطراز M1 $$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 = (\beta_0 + \beta_2) + \beta_1 X_1 \qquad (10.10b)$$

وبالمثل تكون دالة الاستجابة لكل من الطرازين M2 و M3:

آلات الطراز M2 $$E\{Y\} = (\beta_0 + \beta_3) + \beta_1 X_1 \qquad (10.10c)$$

آلات الطراز M3 $$E\{Y\} = (\beta_0 + \beta_4) + \beta_1 X_1 \qquad (10.10d)$$

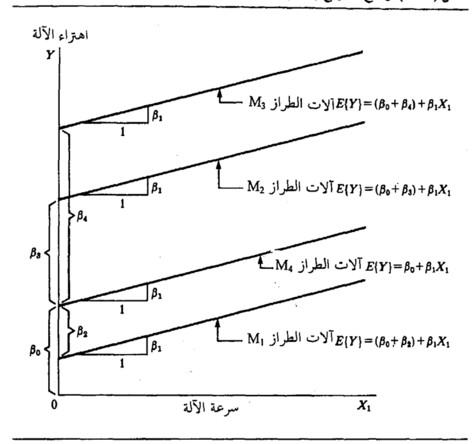
وهكذا تشير دالة الاستجابة (10.10) إلى أن انحدار اهتراء الآلة على سرعتها هو انحدار خطي بالميل نفسه لجميع الطُرُز، وتشير المعاملات β_2، β_3 وβ_4، على الترتيب، إلى أي مدى تكون دالة الاستجابة للطُرُز M1، M2 و$M3$ أعلى (أدنى) منها للطراز M4وذلك لأي مستوى من مستويات سرعة الآلة. ولهذا تقيس β_2، β_3 وβ_4، التأثيرات التفاضلية لصفوف المتغير النوعي على ارتفاع دالة الاستجابة مقاسة دائما بالمقارنة مع دالة الاستجابة للصف الذي تكون فيه $X_2=X_3=X_4 = 0$. وذلك عند أي مستوى معطى لـ X_1. ويوضّح الشكل (١٠-٥) ترتيبا ممكنا لدوال الاستجابة.

وعند استخدام نموذج الانحدار (10.9) قد يود البعض تقدير التأثيرات التفاضلية بمقارنة تختلف عن المقارنة مع الطراز M4 وهذا ممكن بتقدير فروق بين معاملات الانحدار، فعلى سبيل المثال يقيس $\beta_4 - \beta_3$ إلى أي مدى تكون دالة الاستجابة أعلى (أدنى) في آلات الطراز M3 منها في آلات الطراز M2 وذلك لأي مستوى من مستويات سرعة الآلة. ويتأتى هذا بمقارنة (10.10c) و(10.10d) ويكون التقدير النقطي لهذا الفرق هو بالطبع $b_4 - b_1$ والتباين المقدر لهذا التقدير هو:

$$s^2\{b_4 - b_3\} = s^2\{b_4\} + s^2\{b_3\} - 2s\{b_4, b_3\} \qquad (10.11)$$

ويمكن الحصول على التباينات والتغايرات التي تحتاجها مباشرة من مصفوفة التباين والتغاير المقدَّرة لمعالم الانحدار.

شكل (١٠-٥) توضيح للنموذج (10.9) - مثال اهتزاء الآلات



نموذج المرتبة الأولى مع إضافة تفاعلات . إذا كان تأثير التفاعل بين سرعة الآلة وطرازها موجودا في توضيحنا السابق فيمكن تعديل نموذج الانحدار (10.9) كما يلي:

$$Y_i = \beta_0 + \beta_1X_{i1} + \beta_2X_{i2} + \beta_3X_{i3} + \beta_4X_{i4} + \beta_5X_{i1}X_{i2} + \beta_6X_{i1}X_{i3} + \beta_7X_{i1}X_{i4} + \varepsilon_i \quad (10.12)$$

ودالة الاستجابة السابقة لآلات الطراز حيث $X_2 = 0$، $X_3 = 0$ و $X_4 = 0$ هي كما يلي:

$$E\{Y\} = \beta_0 + \beta_1X_1 \quad (10.13a)$$ آلات الطراز M4

وبالمثل فإننا نجد بالنسبة لآلات من الطرز الأخرى:

$$E\{Y\} = (\beta_0 + \beta_2) + (\beta_1 + \beta_5)X_1 \quad (10.13b)$$ آلات الطراز M1

$$E\{Y\} = (\beta_0 + \beta_3) + (\beta_1 + \beta_6)X_1 \quad (10.13c)$$ آلات الطراز M2

M3 آلات الطراز $E\{Y\} = (\beta_0 + \beta_4) + (\beta_1 + \beta_7)X_1$ (10.13d)

وهكذا يتضمن نموذج التفاعل (10.12) أن لكل طراز خط انحداره الخاص بأجزاء مقطوعة مختلفة وميول مختلفة من طراز لآخر.

أكثر من متغير مستقل نوعي واحد

يمكن بسهولة بناء نماذج لحالات تتضمن متغيرين مستقلين أو أكثر كمتغيرات نوعية. لندرس انحدار نفقات الدعاية Y على المبيعات X_1، ونوع (محدودة، غير محدودة)، وجودة إدارة المبيعات (عالية، غير عالية) فيمكن أن نعرف:

$$X_2 = \begin{cases} 1 & \text{إذا كانت الشركة محدودة} \\ 0 & \text{فيما عدا ذلك} \end{cases}$$
$$X_3 = \begin{cases} 1 & \text{إذا كانت الجودة عالية في ادارة المبيعات} \\ 0 & \text{فيما عدا ذلك} \end{cases} \quad (10.14)$$

نموذج من المرتبة الأولى. نموذج انحدار من المرتبة الأولى للمثال السابق هو:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (10.15)$$

ويتضمن هذا النموذج أن دالة الاستجابة لنفقات الدعاية محدورة على المبيعات هي دالة خطية بالميل نفسه لجميع تراكيب نوع الشركة وجودة إدارة المبيعات. وتشير β_2 و β_3 إلى التأثيرات التجميعية التفاضلية على ارتفاع خط الانحدار الناجم عن نوع الشركة وعن جودة إدارة المبيعات، وذلك لأية مستويات معطاة لـ X وللمتغيرات المستقلة الأخرى.

نموذج المرتبة الأولى مع إضافة تفاعلات معينة. نموذج الانحدار من المرتبة الأولى بعد إضافة تأثيرات التفاعل بين أزواج المتغيرات المستقلة هو كما يلي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i2} + \beta_5 X_{i1} X_{i3} + \beta_6 X_{i2} X_{i3} + \varepsilon_i \quad (10.16)$$

لاحظ مضامين هذا النموذج:

دالة الاستجابة	جودة إدارة المبيعات	نوع الشركة
$E\{Y\} = (\beta_0 + \beta_2 + \beta_3 + \beta_6) + (\beta_1 + \beta_4 + \beta_5)X_1$	عالية	محدودة
$E\{Y\} = (\beta_0 + \beta_3) + (\beta_1 + \beta_5)X_1$	عالية	غير محدودة
$E\{Y\} = (\beta_0 + \beta_2) + (\beta_1 + \beta_4)X_1$	منخفضة	محدودة
$E\{Y\} = \beta_0 + \beta_1 X_1$	منخفضة	غير محدودة

لاتختلف جميع دوال الاستجابة للتراكيب المختلفة لنوع الشركة وجودة إدارة المبيعات فقط ولكن التأثيرات التفاضلية لمتغير نوعي واحد على الجزء المقطوع من المحور Y تعتمد على تصنيف المتغير النوعي الآخر، وعلى سبيل المثال عندما ننتقل من (غير محدودة ـ منخفضة الجودة إلى محدودة ـ منخفضة الجودة)، فإن الجزء المقطوع من المحور Y يتغير بمقدار β_2. ولكن لو انتقلنا من محدودة عالية الجودة إلى غير محدودة عالية الجودة فإن الجزء المقطوع يتغير بمقدار $\beta_2 + \beta_6$.

متغيرات مستقلة نوعية فقط

يمكن أيضا بناء نماذج انحدار تتضمن متغيرات مستقلة نوعية فقط. فبالإشارة إلى المثال السابق يمكن دراسة انحدار نفقات الدعاية على نوع الشركة وجودة إدارة المبيعات فقط. وعندئذ يصبح نموذج المرتبة الأولى:

$$Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad (10.17)$$

تعليقات

١ـ يُطلق على النماذج التي تكون جميع المتغيرات المستقلة فيها متغيرات نوعية نماذج تحليل التباين.

٢ـ النماذج التي تحتوي على بعض المتغيرات المستقلة الكمية وبعض المتغيرات المستقلة النوعية، حيث المتغيرات المستقلة الرئيسة ذات الأهمية هي المتغيرات النوعية، أما المتغيرات الكمية فقد أُدخلت لتخفيض تباين حدّ الخطأ، تسمى نماذج تحليل التغاير.

(٤-١٠) المقارنة بين اثنتين أو أكثر من دوال الانحدار

نواجه دائما انحدارات لمجتمعين أو أكثر ونرغب في اختبار أوجه التشابه والاختلاف بينهما ونقدم ثلاثة أمثلة.

١- تشغّل إحدى الشركات خطين لإنتاج قطع الصابون، ودرست العلاقة بين سرعة الخط وكمية النفاية في اليوم. ويقترح رسم الانتشار للبيانات أن علاقة الانحدار بين سرعة الخط وكمية النفاية هي علاقة خطية ولكنها ليست نفسها لكل من خطي الإنتاج. وتبدو الميول وكأنها نفسها تقريبا، أما ارتفاعات خطوط الانحدار فتبدو مختلفة، ونرغب في اختبار ما إذا كان خطا الانحدار متطابقين أم لا، وإذا وُجد أن الخطين ليسا نفسيهما فيجب إجراء دراسة لمعرفة سبب وجود اختلاف في ناتج النفاية.

٢- تدرس إحدى الاقتصاديات العلاقة بين كمية الادخار ومستوى الدخل لمجموعة من الأسر متوسطة الدخل من مناطق حضر ومناطق ريف وذلك باستخدام عينات مستقلة من المجتمعين. ويمكن نمذجة كل من العلاقتين بانحدار خطي. وترغب الاقتصادية في مقارنة ميل أسر الحضر والريف لادخار الكمية نفسها عند المستوى نفسه من الدخل، بمعنى آخر هل خطا لانحدار متشابهان. وإذا لم يكونا متشابهين فهي ترغب في اكتشاف ما إذا كان مقدار الادخار لكل دولار واحد من الدخل متماثلا تماما في المجتمعين. أي ترغب في اكتشاف ما إذا كانت ـ على الأقل ـ كميات الادخار من دولار إضافي في الدخل هي نفسها للمجموعتين وبمعنى آخر هل لخطي الانحدار الميل نفسه.

٣- ابتكرت إحدى الشركات أداتين بالمواصفات نفسها وذلك لقياس الضغط في عملية صناعية. ومن أجل الأداتين قامت بدراسة العلاقة بين قراءة التدريج والضغط الحقيقي كما هو محدد بطريقة مضبوطة تقريبا ولكنها بطيئة ومكلفة. وإذا كان لهما خط الانحدار نفسه فإنه يمكن تطوير جدول معايرة واحد للأداتين وفيما عدا ذلك تحتاج إلى جدولين مختلفين للمعايرة.

عندما يكون افتراض تساوي تباينات حد الخطأ في نماذج الانحدار للمجتمعات المختلفة افتراضا معقولا فيمكن استخدام المتغيرات المؤشرة لاختيار تساوي الدوال المختلفة للانحدار. أما إذا كانت تباينات الخطأ غير متساوية، فمن الممكن جعلها متساوية بصورة تقريبية. على الأقل، من خلال التحويلات.

رأينا فيما سبق أن نماذج الانحدار بمتغيرات مؤشرة، والتي تحتوي على حدود تفاعل، تسمح لنا باختبار تساوي دوال الانحدار للصفوف المختلفة للمتغير النوعي. ويمكن استخدام هذه الطريقة مباشرة لاختيار تساوي دوال الانحدار لمجتمعات مختلفة. وببساطة يمكن اعتبار المجتمعات المختلفة المدروسة كصفوف متغير مستقل، وتعرف متغيرات مؤشرة للمجتمعات المختلفة ثم نطور نموذج انحدار يحتوي على حدود التفاعل المناسبة، وحيث إنه لاتوجد أفكار جديدة في عملية اختبار تساوي دوال الانحدار لمجتمعات مختلفة فسنستخدم مثالين من الأمثلة السابقة لتوضيح الأسلوب.

مثال خطوط إنتاج الصابون

يقدم الجدول (١٠-٦) بيانات عن كمية النفاية Y وسرعة الخط X_1 لمثال خطوط إنتاج الصابون. والمتغير X_2 هو ترميز لخط الإنتاج. ويوضح الشكل (١٠-٦) رسم انتشار للبيانات باستخدام رمزين مختلفين لخطي الإنتاج.

نموذج أولي. قررت المحللة، بناء على رسم الانتشار في الشكل (١٠-٦)، توفيق نموذج الانحدار (10.6) الذي يفترض أن علاقة الانحدار بين كمية النفاية وسرعة الخط هي علاقة خطية لكل من خطي الإنتاج، وأن تباينات حدود الخطأ هي نفسها، ولكنه يسمح بوجود ميلين مختلفين لخطي الانحدار وجزئين مقطوعين مختلفين:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \qquad (10.18)$$

حيث:

$X_{i1} =$ سرعة الخط

$X_{i2} = \begin{cases} 1 & \text{خط الإنتاج رقم ١} \\ 0 & \text{خط الإنتاج رقم ٢} \end{cases}$

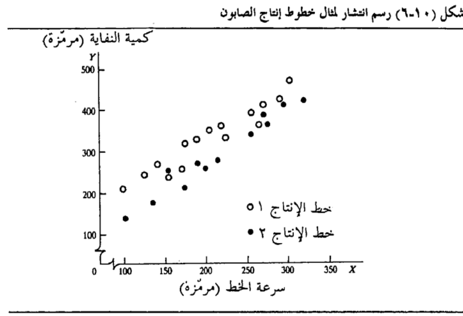
$i = 1, 2, \ldots, 27$

لاحظ أنه من أجل أغراض هذا النموذج، جُمعت المشاهدات الـ 15 من خط الإنتاج 1 والمشاهدات الـ 12 بخط الإنتاج 2 في مجموعة واحدة من 27 مشاهدة.

تشخيصات. يؤدي توفيق بيانات الجدول (١٠-٦) وفقا لنموذج الانحدار (10.18) إلى النتائج المعروضة في الجدول (١٠-٧) وإلى دالة الانحدار التالية:

$$\hat{Y} = 7.57 + 1.322X_1 + 90.39X_2 - 0.1767X_1X_2$$

شكل (١٠-٦) رسم انتشار لمثال خطوط إنتاج الصابون



ويوضح الشكل (١٠-٧) رسما بيانيا للرواسب مقابل $\hat{Y}$ لكل من خطي الإنتاج وهناك رسمان بيانيان لتسهيل تشخيص الاختلافات الممكنة بين خطي الإنتاج. ويتسق كل من الرسمين في الشكل (١٠-٧) مع نموذج الانحدار (10.18) اتساقا معقولا. ويمكن تعليل انقسام الرواسب الموجبة و السالبة إلى 10 و5 على الترتيب، في خط الانتاج 1، وإلى 4 و8 في خط الانتاج 2 إلى عشوائية النتائج. وتدعم الرسوم البيانية للرواسب في مقابل X_2، ورسم الاحتمال الطبيعي للرواسب (غير مبين هنا) صلاحية النموذج الذي تم توفيقه. وبالنسبة لرسم الاحتمال نجد أن معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت فرض التوزيع الطبيعي هو 0.990، وطبقا للجدول (٤-٣) فإن هذا الارتباط عال بما فيه الكفاية كي يشير إلى طبيعية حدود الخطأ.

جدول(١٠-٦) بيانات لمثال خطوط إنتاج الصابون (جميع البيانات مرمزة)

خط الإنتاج ١				خط الإنتاج ٢			
حالة	مقدار النفاية	سرعة الخط		حالة	مقدار النفاية	سرعة الخط	
i	Y_i	X_{i1}	X_{i2}	i	Y_i	X_{i1}	X_{i2}
1	218	100	1	16	140	105	0
2	248	125	1	17	277	215	0
3	360	220	1	18	384	270	0
4	351	205	1	19	341	255	0
5	470	300	1	20	215	175	0
6	394	255	1	21	180	135	0
7	332	225	1	22	260	200	0
8	321	175	1	23	361	275	0
9	410	270	1	24	252	155	0
10	260	170	1	25	422	320	0
11	241	155	1	26	273	190	0
12	331	190	1	27	410	295	0
13	275	140	1				
14	425	290	1				
15	367	265	1				

وترغب المحللة في النهاية إجراء اختيار رسمي لتساوي تباينات حدود الخطأ لكل من خطي الإنتاج، وللحصول على تقديرات مستقلة للتباينات قامت بتوفيق نموذجي انحدار خطيين منفصلين لبيانات كل من خطي الإنتاج وحصلت على النتائج الآتية:

خط الإنتاج	خط الانحدار التوفيقي	MSE	df
1	$\hat{Y}=97.965+1.145X_1$	492.52	13
2	$\hat{Y}=7.574+1.322X_1$	350.13	10

جدول (١٠-٧) نتائج الانحدار لتوفيق نموذج الانحدار (10.18) مثال خطي - إنتاج الصابون

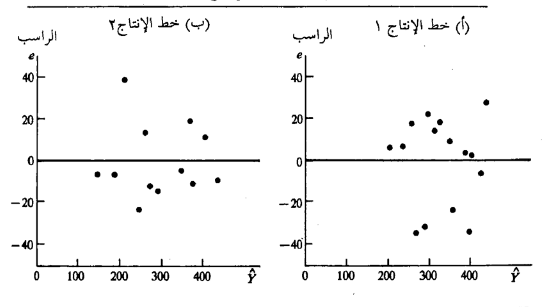
(أ) معاملات الانحدار

معامل الانحدار	معامل الانحدار المقدّر	الانحراف المعياري المقدّر	t^*
β_0	7.57	20.87	.36
β_1	1.322	0.09262	14.27
β_2	90.39	28.35	3.19
β_3	-0.1767	0.1288	-1.37

(ب) تحليل التباين

مصدر التغير	SS	df	MS
معامل الانحدار	169,165	3	56,388
X_1	149,661	1	149,661
$X_2 \mid X_1$	18,694	1	18,694
$X_1X_2 \mid X_1, X_2$	810	1	810
الخطأ	9,904	23	430.6
المجموع	179,069	26	

شكل (١٠-٧) رسومات الرواسب مقابل $\hat{Y}$ لمثال خطي إنتاج الصابون



وباستخدام اختبار F في جدول (١-٤)أ نحصل على:

$$F^* = \frac{492.52}{350.13} = 1.41$$

وعند مستوى معنوية 0.05، نحتاج إلى $F(0.025\ ;\ 13,10) = 0.308$ و $F(0.975;13,10)=3.58$. وحيث إن F^* تقع بين هاتين النهايتين فإننا نستنتج أن لانحداري خطي الإنتاج تباينات خطأ متساوية.

عند هذه النقطة كانت المحللة راضية عن مصداقية نموذج الانحدار (10.18) بحدود خطأ طبيعية، وكانت مستعدة للمضي إلى مقارنة علاقة الانحدار بين كمية النفايات وسرعة الخط، في كل من خطي الإنتاج.

استقراءات حول خطي الانحدار. نختبر تطابق دالتي الانحدار لخطي الإنتاج بدراسة البدائل:

$$H_0 : \beta_2 = \beta_3 = 0 \qquad (10.19)$$

$H_a : \beta_3 = 0$ و $\beta_2 = 0$ ليس كل من

وإحصاءة الاختبار المناسبة معطاة في العلاقة (8.25)

$$F^* = \frac{SSR(X_2, X_1X_2 | X_1)}{2} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4} \qquad (10.19a)$$

حيث تمثل n حجم العينة الكلي للمجتمعين ونجد باستخدام نتائج الانحدار في الجدول (١٠-٧):

$$SSR(X_2, X_1X_2\ |X_1) = SSR(X_2\ |X_1) + SSR(X_1X_2\ |X_1, X_2)$$
$$= 18,694 + 810 = 19,504$$
$$F^* = \frac{19,504}{2} \div \frac{9,904}{23} = 22.65$$

ولضبط α عند مستوى 0.01 نحتاج إلى $F(.99;2,23)=5.67$ وبما أن $F^*=22.65>5.67$ فنستنتج H_a أي أن دالتي الانحدار لخطي الإنتاج ليستا متطابقتين.

اختبرت المحللة بعد ذلك ما إذا كان ميلا خطي الانحدار متساويين وكانت البدائل هنا:

$$H_0 : \beta_3 = 0 \qquad (10.20)$$
$$H_a : \beta_3 \neq 0$$

وإحصاءة الاختبار المناسبة هي إما الإحصاءة t^* في (8.23) أو إحصاءة الاختبار F، الجزئي في(8.22):

$$F^* = \frac{SSR(X_1X_2|X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_1X_2)}{n-4} \qquad (10.20a)$$

وباستخدام نتائج الانحدار في جدول (١٠-٧) وإحصاءة اختبار F الجزئي نحصل على:

$$F^* = \frac{810}{1} \div \frac{9{,}904}{23} = 1.88$$

ومن أجل $\alpha = 0.01$. نحتاج إلى $F(.99; 1{,}23) = 7.88$. وبما أن $F^* = 1.88 \leq 7.88$ فنستنتج H_0، أي أن ميلي دالتي الانحدار متساويان لكل من خطي الإنتاج.

وباستخدام متباينة بونفرّوني (5.3)، يمكن للمحللة إذن أن تستنتج، عند مستوى معنوية عائلي 0.02، أن زيادة معطاة في سرعة الخط تؤدي إلى كمية الزيادة نفسها في النفايات المتوقعة، وذلك في كل من خطي الإنتاج، ولكن الكمية المتوقعة من النفايات المقابلة لسرعة خط معطاة تختلف في الخطين بمقدار ثابت.

ويمكن تقدير هذا الفرق الثابت في خطي الانحدار بالحصول على فترة ثقة لـ β_2. ومن أجل 95 في المائة فترة ثقة نحتاج إلى $t(.975, 23) = 2.069$. وباستخدام النتائج في الجدول(١٠-٧) نحصل على حدود ثقة $90.39 \pm 2.069\ (28.35)$ وبالتالي تكون فترة الثقة لـ β_2:

$$31.7 \leq \beta_2 \leq 149.0$$

وهكذا تستنتج، بمعامل ثقة 95% أن متوسط النفاية لخط الإنتاج ١ تتجاوز، عند أي سرعة خط معطاة، متوسط النفاية لخط الإنتاج ٢ بمقدار يتراوح بين 32 و149.

مثال دراسة معايرة أداة

يعتقد المهندس الذي يقوم بدراسة المعايرة أن دوال الانحدار التي تربط بين قراءة التدريج Y والضغط الفعلي X_1 لكل من الأداتين هي كثيرة حدود من المرتبة الثانية.

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2$$

ولكنهما قد يختلفان من أداة إلى الأخرى، ولذا فإن استخدام النموذج (متخذا متغير انحراف لـ X_1 ليقلل مشاكل الخطية المتعددة. انظر الفصل التاسع):

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \beta_3 X_{i2} + \beta_4 x_{i1} X_{i2} + \beta_5 x_{i1}^2 X_{i2} + \varepsilon_i \qquad (10.21)$$

حيث:

انحراف الضغط الفعلي $= x_{i1} = X_{i1} - \overline{X}_1$

$X_{i2} =$ 1 إذا كانت الأداة B

0 فيما عدا ذلك

لاحظ أن دالة الاستجابة للأداة A حيث $X_2 = 0$، هي:

$$E\{Y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \qquad (10.22a)$$

وتكون دالة الاستجابة للأداة B حيث $X_2 = 1$:

$$E\{Y\} = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) x_1 + (\beta_2 + \beta_5) x_1^2 \qquad (10.22b)$$

وبالتالي فإن اختبار تساوي دالتي الاستجابة يتضمن البدائل:

$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$

H_a: ليست جميع المعالم β_K في H_0، مساوية للصفر (10.23)

وتكون إحصاءة الاختبار المناسبة هي (8.25):

$$F^* = \frac{SSR(X_2, x_1 X_2, x_1^2 X_2 \mid x_1, x_1^2)}{3} \div \frac{SSE(x_1, x_1^2, X_2, x_1 X_2, x_1^2 X_2)}{n-6} \qquad (10.23a)$$

حيث n يمثل حجم العينة المشترك من المجتمعين.

تعليقات

١- يعتبر الأسلوب الموصوف آنفا أسلوبا عاما تماما. وإذا انطوت المسألة على ثلاث أو أكثر من المجتمعات فتضاف ببساطة عندئذ متغيرات مؤشرة إضافية إلى النموذج.

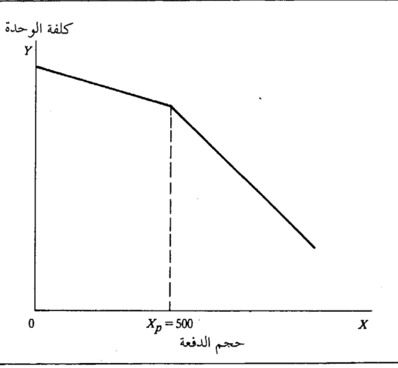
٢- إن استخدام المتغيرات المؤشرة لاختبار تساوي دالتي انحدار أو أكثر مكافىء لأسلوب الاختبار الخطي العام حيث ينطوي توفيق النموذج التام على توفيق انحدارات منفصلة لبيانات كل مجتمع على حدة وينطوي توفيق النموذج المخفّض على توفيق انحدار واحد للبيانات جميعها معتبرة كبيان واحد.

(١٠ـ٥) استخدامات أخرى للمتغيرات المؤشرة

الانحدار الخطي قطعة فقطعة

يتبع انحدار Y على X في بعض الأحيان علاقة خطية بالذات ضمن مدى معين للمتغير X، ولكنه يتبع علاقة خطية مختلفة خارج ذلك المدى. وعلى سبيل المثال قد يتبع انحدار تكلفة وحدة إنتاج Y على حجم دفعة الانتاج علاقة انحدار خطية معينة حتى$X_p = 500$ وعند هذه النقطة، يتغير الميل بسبب فعاليات في عملية التشغيل لاتكون ممكنة إلا عندما يكون حجم الدفعة أكبر من 500. وعلى سبيل المثال، قد يكون هناك هبوط شديد في سعر الوحدة عند شراء المواد الأولية لدفعات إنتاج تتجاوز الـ 500 ويوضّح الشكل (١٠ـ٨) هذه الحالة.

شكل (١٠ـ٨) توضيح للانحدار الخطي قطعة فقطعة



ونستعرض الآن كيفية استخدام المتغيرات المؤشرة لتوفيق انحدارات خطية قطعة فقطعة ومؤلفة من قطعتين. وسنتابع الحالة التي تكون فيها X_p (النقطة التي يتغير عندها الميل) معروفة.

وبالعودة إلى توضيح حجم الدفعة، حيث نعلم أن الميـل يتغـير عنـد $X_p = 500$، يمكن التعبير عن النموذج لهذا المثال كما يلي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 (X_{i1} - 500) X_{i2} + \varepsilon_i \qquad (10.24)$$

حيث:

$X_{i1} =$ حجم الدفعة

$$X_{i2} = \begin{cases} 1 & X_{i1} > 500 \text{ إذا كان } \\ 0 & \text{فيما عدا ذلك} \end{cases}$$

وللتحقق من أن نموذج الانحدار (10.24) يقدم انحدارا خطيا بقطعتين، لنأخذ دالة الاستجابة لهذا النموذج:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 (X_1 - 500) X_2 \qquad (10.25)$$

وعندما يكون $X_1 \le 500$ يكون $X_2 = 0$ وبالتالي تصبح (10.25):

$$E\{Y\} = \beta_0 + \beta_1 X_1 \qquad X_1 \le 500 \qquad (10.25a)$$

وفي المقابل عندما يكون $X_1 > 500$ فإن $X_2 = 1$ ونحصل على:

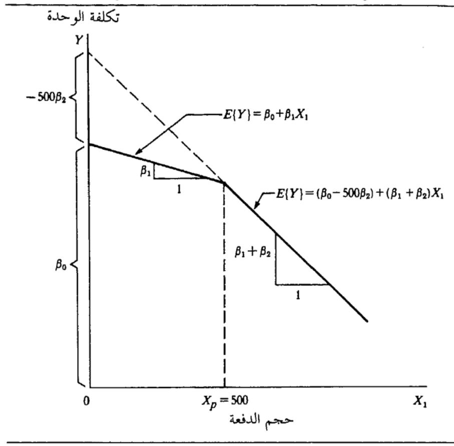
$$E\{Y\} = (\beta_0 - 500\beta_2) + (\beta_1 + \beta_2) X_1 \qquad X_1 > 500 \qquad (10.25b)$$

وهكذا يكون ميـلا خطـي الانحـدار β_1 و $\beta_2 + \beta_1$، ويكـون الجـزءان المقطوعـان، β_0 و $\beta_0 - 500\,\beta_2$. ويوضح الشكل (١٠-٩) هذه المعالم، والسبب في طرح β_2 00 من β_0 هو أن β_2تتعلق بزيادة مقدارها الواحد في X_1، ولكننـا نقـوم هنـا بقيـاس وقـع التأثـير التفاضلي في الميل في الاتجاه السالب الممتد من $X_p = 500$ إلى الصفر.

مثال. يحتوي الجدول (١٠-٨)أ على ثمانية مشاهدات حـول تكـاليف الوحـدة لحجـوم دفعات معينة. ومن المعروف أن ميـل دالـة الاسـتجابة يتغـير عنـد $X_p = 500$ ولذلـك سنستخدم النموذج (10.24).

ويحتوي الجدول (١٠-٨)ب على مصفوفة **X** لمثالنا، والعمود الأيسر لـ **X** هـو عمـود من الأرقام 1 كالمعتاد. ويحتوي العمود الذي يليـه القيـم X_{i1}. ويحتـوي العمـود الأخـير على اليمـين $(X_{i1} - 500)X_{i2}$، ويتـألف مـن القيـم 0 لجميـع حجـوم الدفعـات حتـى 500، و$X_{i1} - 500$ لحجوم الدفعات التي تتجاوز 500.

شكل (١٠-٩) توضيح لمعنى معاملات الانحدار في دالة الاستجابة (10.25) وهي دالة قطعة فقطعة.



جدول (١٠-٨) بيانات ومصفوفة X لانحدار خطي قطعة فقطعة – مثال حجم الدفعة

(أ)

الدفعة i	تكلفة الوحدة Y_i (بالدولار)	حجم الدفعة X_i
1	2.57	650
2	4.40	340
3	4.52	400
4	1.39	800
5	4.75	300
6	3.55	570
7	2.49	720
8	3.77	480

(ب)

$$\mathbf{X} = \begin{bmatrix} 1 & 650 & 150 \\ 1 & 340 & 0 \\ 1 & 400 & 0 \\ 1 & 800 & 300 \\ 1 & 300 & 0 \\ 1 & 570 & 70 \\ 1 & 720 & 220 \\ 1 & 480 & 0 \end{bmatrix} \quad \begin{matrix} \text{columns: } X_0, \; X_1, \; (X_1 - 500)X_2 \end{matrix}$$

ويصبح توفيق نموذج الانحدار (10.24)، عند هذه النقطة مجرد روتين. ودالة الاستجابة التوفيقية هي:

$$\hat{Y} = 5.89545 - 0.00395X_1 - 0.00389(X_1 - 500)X_2$$

ونقدِّر من دالة الاستجابة التوفيقية هذه أن التكلفة المتوقعة للوحدة ستهبط بمقدار 0.00395 لكل زيادة بمقدار واحد في حجم الدفعة، وذلك عندما يكون حجم الدفعة أقل من 500، ويهبط بمقدار $0.00395 + 0.00389 = 0.00784$ عندما يكون حجم الدفعة أكبر من أو يساوي 500.

ملاحظة

من السهل تعميم نموذج الانحدار (10.24) إلى خطوط انحدار بأكثر من قطعتين. فعلى سبيل المثال، إذا كان ميل خط الانحدار في مثال حجم الدفعة السابق، يتغير في الواقع عند كل من $X = 500$ و$X = 800$ فسيصبح النموذج :

$$Y_i = \beta_0 + \beta_1X_{i1} + \beta_2(X_{i1} - 500)X_{i2} + \beta_3(X_{i1} - 800)X_{i3} + \varepsilon_i \qquad (10.26)$$

حيث: $X_{i1} =$ حجم الدفعة

$X_{i2} =$ 1 إذا كان $X_{i1} > 500$
0 فيما عدا ذلك

$X_{i3} =$ 1 إذا كان $X_{i1} > 800$
0 فيما عدا ذلك

انقطاع في دالة الانحدار

وأحيانا لاتغير دالة انحدار خطية ميلها عند قيمة ما X فقط، ولكن قد يكون لها أيضا قفزة عند هذه القيمة. ويوضح الشكل (١٠-١٠) هذه الحالة. ويجب الآن إدخال متغير مؤشر آخر ليُعنى بهذه القفزة. افترض أنه يُراد حدر الوقت اللازم لحل مشكلة بنجاح Y على درجة تعقيد تلك المشكلة X مقيسة بمتغيرات كمية من 0 إلى 100. وأنه من المعروف أن خط الاستجابة يغير ميله عند $X_p = 40$ وأنه يُعتقد أن علاقة الانحدار قد تكون غير مستمرة في هذه النقطة فعندئذ نضع نموذج الانحدار:

$$Y_i = \beta_0 + \beta_1X_{i1} + \beta_2(X_{i1} - 40)X_{i2} + \beta_3X_{i3} + \varepsilon_i \qquad (10.27)$$

حيث:

درجة تعقيد المشكلة = X_{i1}

$$X_{i2} = \begin{cases} 1 & \text{إذا كان } X_{i1} > 40 \\ 0 & \text{فيما عدا ذلك} \end{cases}$$

$$X_{i3} = \begin{cases} 1 & \text{إذا كان } X_{i1} > 40 \\ 0 & \text{فيما عدا ذلك} \end{cases}$$

وتكون دالة الاستجابة لنموذج الانحدار (10.27)

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 (X_1 - 40) X_2 + \beta_3 X_3 \qquad (10.28)$$

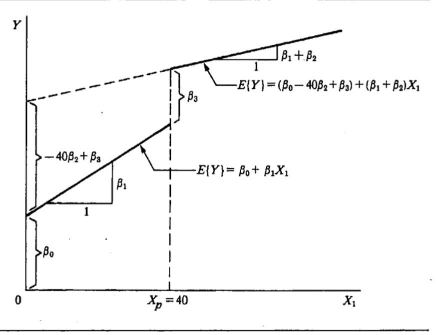
وعندما يكون $X_1 \leq 40$ فإن $X_2 = 0$ و $X_3 = 0$ ولهذا تصبح (10.28):

$$E\{Y\} = \beta_0 + \beta_1 X_1 \qquad X_1 \leq 40 \qquad (1.28a)$$

وبالمثل، عندما يكون $X_1 > 40$ فإن $X_2 = 1$ و $X_3 = 1$ ولهذا تصبح النموذج (10.28)

$$E\{Y\} = (\beta_0 + 40\beta_2 + \beta_3) + (\beta_1 + \beta_2) X_1 \qquad X_1 > 40 \qquad (10.28b)$$

شكل (١٠-١٠) توضيح لدالة الاستجابة (10.28) حيث الانحدار الخطي غير مستمر وقطعة فقطعة.



ودالتا الاستجابة هاتان مبينتان في الشكل (١٠ـ١٠) مع المعالم التي تنطوي عليها الدالتان. لاحظ أن β_3 تمثل الفرق بين متوسطي الاستجابة لخطي الانحدار عند $X_p = 40$، ويمثل β_2 الفرق بين الميلين.

ولايشكل تقدير معاملات الانحدار للنموذج (10.27) مشكلة جديدة. ويمكن اختبار ما إذا كان $\beta_3 = 0$ أم لا بالطريقة المعتادة. وإذا استنتجنا أن $\beta_3 = 0$ فتكون دالة الانحدار مستمرة عند X_p بحيث ينطبق النموذج السابق للانحدار الخطي قطعة فقطعة.

تطبيقات السلاسل الزمنية

كثيرا مايستخدم الاقتصاديون والمحللون التجاريون بيانات السلاسل الزمنية في تحليل الانحدار. فعلى سبيل المثال يمكن حدر الادخارات Y على الدخل X حيث بيانات الادخار والدخل تتعلق بعدة سنوات. ويمكن أن يكون النموذج المقترح:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \qquad t = 1,\ldots,n \tag{10.29}$$

حيث Y_t و X_t الادخار والدخل عند الزمن t، على الترتيب. افترض أن الفترة الزمنية التي تغطيها البيانات تشتمل على سنوات حرب وسنوات سلام، وأن هذا العامل ينبغي أخذه في الاعتبار، إذ يُتوقع أن يتجه الادخار خلال سنوات الحرب إلى أن يكون أعلى. وبالتالي يمكن أن يكون النموذج التالي مناسبا:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \varepsilon_t \tag{10.30}$$

حيث:

$X_{t1} =$ الدخل

$$X_{t2} = \begin{cases} 1 & \text{إذا كان الفترة } t \text{ زمن سلام} \\ 0 & \text{فيما عدا ذلك} \end{cases}$$

لاحظ أن نموذج الانحدار (10.30) يفترض أن النزوع الهامشي للتوفير β_1 يبقى ثابتا في سنوات الحرب وسنوات السلام. وأن ارتفاع منحني الاستجابة فقط هو الذي يتأثر بهذا المتغير النوعي.

تظهر إحدى الاستخدامات الأخرى للمتغيرات المؤشرة في تطبيقات السلاسل الزمنية عندما نستخدم بيانات شهرية أو ربعية (ربع سنوية). افترض أن المبيعات

الربعية Y انحدرت على نفقات الإعلان الربعية X_1 وعلى الدخل الشخصي القابل للإنفاق في الربع X_2. إذا كان للتأثيرات الموسمية أثر على المبيعات الربعية فسيكون نموذج الانحدار من المرتبة الأولى المستوعب للتأثيرات الموسمية كما يلي:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \beta_4 X_{t4} + \beta_5 X_{t5} + \varepsilon_t \qquad (10.31)$$

حيث:

$X_{t1} =$ نفقات الإعلان الربعية

$X_{t2} =$ الدخل القابل للإنفاق في الربع

$X_{t3} =$ 1 إذا كنا في الربع الأول، 0 فيما عدا ذلك

$X_{t4} =$ 1 إذا كنا في الربع الثاني، 0 فيما عدا ذلك

$X_{t5} =$ 1 إذا كنا في الربع الثالث، 0 فيما عدا ذلك

(١٠-٦) بعض الاعتبارات في استخدام المتغيرات المؤشرة المستخدمة

المتغيرات المؤشرة في مقابل رموز مخصصة

أحد البدائل لاستخدام المتغيرات المؤشرة المستقلة هو استخدام رموز نقوم بتخصيصها. اعتبر، على سبيل المثال، المتغير المستقل " تواتر استخدام مُنتَج ". وله ثلاثة فصول: مستخدم بكثرة، مستخدم من وقت لآخر، غير مُستخدم. ومع أسلوب الرموز المخصصة نستخدم متغيرا مستقلا واحدا ونخصص قيما للفصول المختلفة؛ مثلا:

الفصل	X_1
مستخدم بكثرة	3
مُستخدم من وقت لآخر	2
غير مُستخدم	1

والرموز المخصصة هي، بـالطبع، اختياريـة ويمكـن أن تكـون أي مجموعـة أخـرى مـن الأعداد ويصبح النموذج لهذا المثال مع رموز مخصصة وافـتراض عـدم وجـود متغـيرات مستقلة أخرى:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \qquad (10.32)$$

والصعوبة الأساسية في استخدام رموز مخصصة هي أنها تُعـرّف مقياسـا لفصـول المتغير النوعي قد تكون غير منطقية. ولرؤية ذلك بصـورة ملموسـة، اعتـبر متوسـطات الاستجابة لنموذج الانحدار (10.32) وذلك من أجل الفصول الثلاثة للمتغير النوعي.

الصف	$E\{Y\}$
مستخدم بكثرة	$E\{Y\} = \beta_0 + \beta_1(3) = \beta_0 + 3\beta_1$
مُستخدم من وقت لآخر	$E\{Y\} = \beta_0 + \beta_1(2) = \beta_0 + 2\beta_1$
غير مُستخدم	$E\{Y\} = \beta_0 + \beta_1(1) = \beta_0 + \beta_1$

لاحظ المقتضى الرئيس لهذا :

$E\,(Y|$ مستخدم بكثرة$)$ - $E\,(Y|$ مستخدم من وقت لآخر$)$

$= E(Y|$ مستخدم من وقت لآخر$)$ - $E\,(Y|$ غير مستخدم$)$ = β_1

وهكذا يتضمن الترميز 1، 2 و3 أن متوسط الاستجابة يتغير بالمقدار نفسه عندما نمضي من "غير مستخدم" إلى "مستخدم من وقت لآخر" كتغيره عند المضـي مـن "مسـتخدم من وقت لآخر" إلى "مستخدم بكثرة". وقد لايتفق هذا مع الواقـع الفعلـي، وإنمـا هـو نتيجـة للـترميز 1، 2 و3 الـذي يخصـص مسـافات متســاوية بـين الصفـوف الثلاثـة للمستخدمين. وبالطبع قد يتضمن تخصيص آخر لـلرموز مسـافات فاصلـة مختلفـة بـين فصول المتغير النوعي، إلا أن ذلك سيبقى في العادة اختياريا.

وفي المقابل فإن المتغيرات المؤشرة لاتضع افتراضات للمسافات بين الفصــول، بـل تعتمد على البيانات لتبيان التأثيرات التفاضلية الحاصلـة. وإذا اسـتخدمنا للمثـال نفسـه متغيرين مؤشرين X_1 وX_2 مثلا، لتمثيل المتغير النوعي وذلك كما يلي:

الفصل	X_1	X_2
مستخدم بكثرة	1	0
مُستخدم من وقت لآخر	0	1
غير مُستخدم	0	0

فسيكون نموذج الانحدار:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad (10.33)$$

وتقيس β_1 هنا التأثير التفاضلي:

$E\{Y|$مستخدم بكثرة$\}$ - $E\{Y|$غير مستخدم$\}$

وتقيس β_2:

$E\{Y|$مستخدم من وقت لآخر$\}$ - $E\{Y|$غير مستخدم$\}$

وهكذا تقيس β_2 التأثير التفاضلي بين المستخدم من وقت لآخر وغير المستخدم، ويقيس $\beta_2 - \beta_1$ التأثير التفاضلي بين المستخدم بكثرة والمستخدم من وقت لآخر. ويُلاحظ أنه ليس هناك قيود اختيارية يُراد لهذه التأثيرات التفاضلية أن تحققها.

المتغيرات المؤشرة في مقابل المتغيرات الكمية

إذا كان أحد المتغيرات المستقلة كميا كالعمر فيمكن، مع ذلك، استخدام متغير مؤشر بدلا منه. إذ يمكن تحويل المتغير الكمّي، العمر، على سبيل المثال، بتجميع الأعمار في صفوف مثل تحت 21، 34 – 21، 49 – 35،... إلخ. ويمكن عندئذ استخدام متغيرات مؤشرة لفصول هذا المتغير النوعي الجديد. ويثير هذا الأسلوب للوهلة الأولى التساؤل بسبب ضياع معلومات عن الأعمار الفعلية، وأكثر من ذلك أُضيفت معالم إضافية إلى النموذج مما يؤدي إلى تقليل درجات الحرية الموافقة لـ *MSE*.

وعلى الرغم من ذلك فهناك حالات يكون فيها استبدال متغير مؤشر بمتغير كمي استبدالا مناسبا. اعتبر، على سبيل المثال، بحثا ميدانيا كبيرا لدراسة العلاقة بين الممتلكات السائلة Y والعمر X لمسؤول الأسرة. وتشمل الدراسة ألفين من الأسر ولهذا فإن فقد 10 إلى 20 درجة حرية غير ذي بال. ويرتاب المحلل بشدة حول شكل دالة الانحدار التي قد تكون معقدة جدا ولذا فقد يفضّل أسلوب المتغير المؤشر كي يحصل على معلومات حول الشكل دون وضع أية افتراضات تتعلق بالصيغة الدالية لعلاقة الانحدار. ويُتاح للمحلل الذي يرتاب في الصيغة الدالية لعلاقة انحدار، قد تكون معقدة، بديل آخر ينتفع أيضا بالمتغيرات المؤشرة. إذ يمكنه استخدام المتغير الكمي، العمر، ولكن بعلاقة انحدار قطعة فقطعة مع عدد من القطع. ومرة أخرى يخفض هذا

الأسلوب من درجات الحرية المخصصة لتقدير MSE، ولكن هذا لايهـم في الدراسـات ذات الحجم الكبير. وستكون الفائدة في الحصـول على معلومـات حـول نـوع علاقـة الانحدار دون وضع افتراضات قوية حول صيغتها الدالية.

ترميزات أخرى للمتغيرات المؤشرة

كما سبق أن ذكرنا هناك ترميزات كثيرة ممكنة للمتغيرات المؤشرة وسنذكر الآن بديلين للترميز 0 و1 مع $c - 1$ متغيرا مؤشرا لمتغير نوعي له c من الفصول.

في مثال خطة التأمين المبتكرة حيث Y الوقت المنصرم حتى تبنّي الابتكـار، و X_1 حجم شركة التأمين، والمتغير المستقل الثاني هو نوع الشركة (مساهمة ـ تعاونية) يمكن استخدام الترميز التالي:

$$X_2 = \begin{cases} 1 & \text{شركة مساهمة} \\ -1 & \text{شركة تعاونية} \end{cases} \qquad (10.34)$$

ويكون نموذج الانحدار الخطي من المرتبة الأولى، في هذه الحالة:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad (10.35)$$

حيث دالة الاستجابة:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (10.36)$$

وستكون دالة الاستجابة كالتالي لكل من نوعي الشركات:

$$E\{Y\} = (\beta_0 + \beta_2) + \beta_1 X_1 \qquad (10.36a) \qquad \text{شركة مساهمة}$$

$$E\{Y\} = (\beta_0 - \beta_2) + \beta_1 X_1 \qquad (10.36b) \qquad \text{شركة تعاونية}$$

ولهذا فإنه يمكن النظر إلى β_0 على أنها "متوسط" الجـزء المقطـوع مـن خـط الانحـدار، والذي يختلف عن الجزء المقطوع لكل مـن الشـركات المسـاهمة والشـركات التعاونيـة بمقدار β_2 في اتجاهين مختلفين ـ وينطوي اختبار ما إذا كان لكل من نوعـي الشـركات خط الانحدار نفسه أم لا على $H_0 : \beta_2 = 0$ و $H_a : \beta_2 \neq 0$.

والبديل الثاني لخطة الترميز هو استخدام 0 و1 متغيرات مؤشرة لكل مـن c صفـا للمتغير النوعي ثم حذف الحد المتعلق بـالجزء المقطـوع في نمـوذج الانحـدار. وفي مثـال خطة التأمين المبتكرة يمكن أن نكتب:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad (10.37)$$

حيث:

X_{i1} = حجم الشركة

X_{i2} = 1 شركة تعاونية
0 فيما عدا ذلك

X_{i3} = 1 شركة مساهمة
0 فيما عدا ذلك

وستكون دالتا الاستجابة هنا:

$$E\{Y\} = \beta_2 + \beta_1 X_1 \quad (10.38a)$$ شركة مساهمة

$$E\{Y\} = \beta_3 + \beta_1 X_1 \quad (10.38b)$$ شركة تعاونية

وسينطوي اختبار ما إذا كان خطا الانحدار متطابقين على البدائل: $H_0: \beta_2 = \beta_3$ و $H_a: \beta_2 \neq \beta_3$. وقد نوقش هذا النوع من الاختبار في الفقرة (٨ -٢).

مسائل

(١٠-١) استاء أحد الطلبة ممن استخدموا نموذج انحدار يشمل متغيرات مؤشرة عندما وجد أن المخرج على مطبوعة حاسب اقتصر على:

XTRANSPOSE X SINGULAR ماعساه أن يكون مصدر الصعوبة؟.

(١٠-٢) بالإشارة إلى نموذج الانحـدار (10.4) ارسم بيانيا منحنيات الاستجابة لهذا النموذج إذا كان $\beta_0 = 25.3$، $\beta_1 = 0.20$ و $\beta_2 = -12.1$.

(١٠-٣) في دراسة انحدار للعوامل التي تؤثر في زمن التعلم لعمل معين (مقاسا بالدقائق). اعتُبر جنس المتعلم كمتغير مستقل (X_2) وتمّ ترميزه على الشكل التالي $X_2 = 1$ إذا كان المتعلم ذكرا و $X_2 = 0$ إذا كان المتعلم أنثى. وقد وُجد أن $b_2 = 22.3$ و $s\{b_2\} = 3.8$. وتساءل أحد الملاحظين عما إذا كان نظام الترميز للجنس عادلا. ذلك لأن الدراسة أنتجت معاملا موجبا مما يوحي أن أزمنة التعلم للذكور أطول منها للإناث. علّق.

(١٠-٤) بالإشـارة إلى المسألة (٢-١٨) **صيانة الآلات الحاسبة**. مستخدمو الآلات الحاسبة المكتبية هم إما معاهد تدريب تستخدم نموذج الطلاب، أو شـركات أعمال تستخدم النموذج التجاري. ويرغب أحد المحللين في ترايسيتي توفيق نموذج انحدار يشتمل على متغيرات مستقلة هي عدد الآلات الحاسبة التي تتطلب صيانة (X_1) ونوع الآلة الحاسبة (X_2) ويقدر تأثير نموذج الآلة الحاسبة (S ترمز لطلاب وC ترمز لتجاري) على عـدد الدقائق التي يستغرقها أداء خدمة مطلوبة. أوضحت السجلات أن النماذج التي طلبت صيانة في 18 طلبا كانت كالآتي:

i :	1	2	3	4	5	6	7	8	9
النموذج:	C	S	C	C	C	S	S	C	C

i :	10	11	12	13	14	15	16	17	18
النموذج:	C	C	S	S	C	S	C	C	C

افترض أن نموذج الانحدار (١٠-٤) مناسب وليكن $X_2 = 1$ لنموذج "طـلاب" و$X_2 = 0$ لنموذج " تجاري".

أ - اشرح معنى جميع معاملات الانحدار في النموذج.

ب - وفق نموذج الانحدار واكتب دالة الانحدار المقدرة.

جـ - قدر تأثير نموذج الآلة الحاسبة على متوسط وقت الخدمة مستخدما فترة ثقة بمعامل 0.95 فسر معنى الفترة التي حصلت عليها.

د - لماذا يرغب المحلل في إدخال عدد الآلات الحاسبة كمتغير في نموذج الانحدار في حين أنه يهتم بتقدير تأثير نوع الآلة الحاسبة على وقت الخدمة؟

هـ - احسب الرواسب وارسمها بيانيا في مقابل X_1X_2. هل هناك مايشير إلى أن وجود حد للتفاعل في النموذج قد يكون مفيدا ؟

(١٠-٥) بالإشارة إلى المسألة (٢-١٧) **المعدل التراكمي**. لدى أحد مساعدي مدير القبول حدس بأن القدرة التنبؤية للنموذج قد تتحسن بإضافة معلومات توضح ما إذا كان الطالب قد اختار تخصصا رئيسا عند تقديمه لطلب القبول مفترضا أن نموذج الانحدار (10.4) مناسب حيث X_1 درجة امتحان الدخول، $X_2 = 1$ إذا اختار الطالب تخصصا رئيسا عند تقديمه لطلب القبول، $X_2 = 0$ إذا لم يكن التخصص الرئيس للطالب محددا. وقد اختار الطلاب 3، 4، 5، 6، 9 و17 تخصصا رئيسا عند تقديم الطلب.

أ - اشرح كيف يمكن تفسير معاملات الانحدار في النموذج (10.4) في هذه الحالة.

ب - قم بتوفيق نموذج الانحدار، ثم اكتب دالة الانحدار المقدَّرة.

جـ - اختبر ما إذا كان ممكنا إسقاط المتغير X_2 من النموذج. استخدم $\alpha = 0.01$ واكتب البدائل وقاعدة القرار، والنتيجة.

د - أوجد الرواسب للنموذج (10.4)، ثم ارسمها بيانيا في مقابل X_1X_2. هل هناك أي دليل من رسمك يساعد على القول إن وجود حد تفاعل في النموذج قد يكون مفيدا؟

(١٠-٦) بالإشارة إلى نموذجي الانحدار (10.4) و(10.6). هل للنتيجة $\beta_2 = 0$ المعنى نفسه في كل من النموذجين؟ اشرح.

(١٠-٧) بالإشارة إلى نموذج الانحدار (10.6) ارسم بيانيا منحنيات الاستجابة لهذا النموذج إذا كان $\beta_0 = 25$، $\beta_1 = 0.30$، $\beta_2 = -12.5$ و$\beta_3 = 0.05$. صف طبيعة تأثير التفاعل.

(١٠-٨) بالإشارة إلى المسألتين (٢-١٨) و(١٠-٤) **صيانة آلات حاسبة**

أ - قم بتوفيق نموذج الانحدار (10.6) واكتب دالة الانحدار المقدَّرة.

ب - اختبر ما إذا كان يمكن إسقاط حد التفاعل من النموذج مستخدما

$\alpha = 0.1$. اكتب البدائل وقاعدة القرار والنتيجة ما هي القيمة -P للاختبار؟ وإذا لم يكن من الممكن إسقاط حد التفاعل من النموذج صف طبيعة تأثيره.

(١٠-٩) بالإشارة إلى المسألتين (٢-١٧) و(١٠-٥). **المعدل التراكمي.**

أ - قم بتوفيق نموذج الانحدار (10.6) اكتب دالة الانحدار المقدَّرة.

ب - اختبر ما إذا كان يمكن إسقاط حد التفاعل مستخدما $\alpha = 0.05$. اكتب البدائل وقاعدة القرار والنتيجة. وإذا لم يكن ممكنا إسقاط حد التفاعل صف طبيعة تأثيره.

(١٠-١٠) في تحليل انحدار إصابات الرأس أثناء العمل لعمال مستودع بسبب سقوط بعض الأشياء، Y قياس لشدة الإصابة، X_1 مؤشر يعكس وزن الشيء ومسافة السقوط، X_2 وX_3 متغيرات مؤشرة لطبيعة الوقاية التي تغطي الرأس عند وقوع الحادثة ومرمّزة كالآتي:

نوع الحماية	X_2	X_3
قبعة صلبة	1	0
قبعة للصدمات	0	1
لاشيء	0	0

ودالة الاستجابة المستخدمة في هذه الدراسة هي:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

أ - استنتج دالة الاستجابة لكل نوع من أنواع الوقاية.

ب - لكل من الأسئلة التالية حدد البدائل H_0 و H_a للاختبار المناسب:

(١) في حالة ثبات X_1 هل ارتداء قبعة للصدمات يقلل من شدة الإصابة المتوقعة بالمقارنة مع عدم ارتداء أي شيء؟ (٢) في حالة ثبات X_1 هل تكون شدة الإصابة المتوقعة نفسها في حالة إرتداء قبعة صلبة وحالة ارتداء قبعة للصدمات؟

(١٠-١١) بالإشارة إلى نموذج انحدار اهتراء آلة (10.12)، إفترض أن المتغيرات المؤشرة قد عُرفت كالآتي: $X_2 = 1$ إذا كان طراز الآلة M2 وصفرا فيما عدا ذلك،

$X_3 = 1$ إذا كان طراز الآلة M3 وصفرا فيما عدا ذلك، $X_4 = 1$ إذا كان طراز الآلة M4 وصفرا فيما عدا ذلك. وضّح معنى كل من (١) β_3، (٢) $\beta_4 - \beta_3$، (٣) β_1 و(٤) β_7.

(١٠-١٢) بالإشارة إلى نموذج انحدار نفقات الدعاية (10.16)

أ - كيف يمكن تفسير β_4؟ ما معنى $\beta_3 - \beta_2$ هنا؟

ب - اكتب البدائل لاختبار ما إذا كانت دوال الاستجابة نفسها في الشركات المحدودة وغير المحدودة، مع وجود إدارة مبيعات عالية الجودة.

(١٠-١٣) استخدم أحد المتدربين في أبحاث التسويق في المكتب الوطني لسلسلة من محلات الأحذية دالة الاستجابة التالية لدراسة التأثيرات الموسمية (شتاء - ربيع - صيف - خريف) على مبيعات صنف معين من الأحذية:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

والمتغيرات المؤشرة X معرفة كالآتي $X_1 = 1$ إذا كان الوقت شتاء وصفرا فيما عدا ذلك، $X_2 = 1$ إذا كان الوقت ربيعا وصفرا فيما عدا ذلك، $X_3 = 1$ إذا كان الوقت خريفا وصفرا فيما عدا ذلك. وبعد توفيق النموذج، قام باختبار معاملات الانحدار β_k ($k = 0, 1, 2, 3,$) وحصل على النتائج الآتية حيث $\alpha = .05$. لكل اختبار $\beta_0 \neq 0$، $\beta_1 = 0$، $\beta_2 \neq 0$ و$\beta_3 \neq 0$.

وجاء في تقريره مايلي "يبين تحليل الانحدار أن العوامل المناخية والموسمية لاتؤثر في مبيعات هذا الصنف من الأحذية في الشتاء والتأثيرات الموسمية موجودة في الفصول الأخرى. هل توافق على هذا التفسير لنتائج الاختبار؟ ناقش.

(١٠-١٤) **تثمين الممتلكات**: درس أحد مستشاري الضرائب العلاقة القائمة بين سعر البيع المعروض والقيمة التقديرية في مصلحة الضرائب لمبنى سكن لأسرة واحدة في منطقة ضريبية كبيرة وحصل على بيانات عينة عشوائية من تسع مبيعات حديثة لمبنى سكن أسرة واحدة تقع على تقاطع شارعين وحصل أيضا على بيانات عينة عشوائية من 14 مبيعا لمبنى سكن أسرة واحدة لايقع في التقاطعات. وتعرض البيانات التالية كلا من سعر المبيع Y والقيمة

التقديرية X_1 معطاة بآلاف الدولارات افترض أن التباينات متساوية في المجتمعين وأن نموذج الانحدار (10.6) مناسب.

مباني تقع عند تقاطع

i	1	2	3	4	5	6	7	8	9
X_{i1}	17.5	12.5	20.0	16.0	15.0	14.7	17.5	12.3	11.5
Y_i	56.2	42.5	68.6	54.8	50.0	47.5	56.9	34.0	39.0

مباني لا تقع عند تقاطع

i	1	2	3	4	5	6	7
X_{i1}	10.0	13.8	15.0	19.5	17.0	12.5	14.5
Y_i	31.2	36.9	41.0	51.8	48.0	33.3	38.0

i	8	9	10	11	12	13	14
X_{i1}	12.8	12.0	16.0	10.0	17.0	10.8	15.0
Y_i	35.9	32.0	44.3	29.0	46.1	30.0	42.0

أ ـ ارسم بيانات العينة لكل من المجتمعين في شكل واحد مستخدما رموزا مختلفة لكل من العيّنتين. هل تبدو علاقة الانحدار نفسها لكل من المجتمعين؟

ب ـ اختبر تطابق دالتي الانحدار لكل من مباني التقاطع والمباني التي لاتقع عند تقاطع. اضبط مخاطرة الخطأ من النوع الأول عند 0.10 اكتب البدائل وقاعدة القرار والنتيجة.

جـ ـ ارسم دالتي الانحدار المقدّرتين للمجتمعين وصف طبيعة الاختلاف بينهما.

د ـ قدِّر باستخدام فترات الثقة الفرق بين ميلي دالتي الانحدار مستخدما معامل ثقة 0.90.

هـ ـ احسب الرواسب لكل عينة وارسمها في مقابل $\hat{Y}$ لكل من العيّنتين بصورة منفصلة. هل تبدو فرضية تساوي تباينات الخطأ مقبولة هنا؟

و ـ مستخدما اختبار F اختبر تساوي تباينات الخطأ لكل نوع من المواقع استخدم α =0.05. اكتب البدائل وقاعدة القرار والنتيجة. ماهي القيمة -P لهذا الاختبار؟

ز ـ ارسم بيانيا في شكلين منفصلين الرواسب في مقابل X_1، X_2 وX_2 X_1 .

وقم أيضا بإعداد رسم احتمال طبيعي للرواسب. فسّر رسومك. ولخّص نتائجك.

(١٠-١٥) **اختبار عجلة سيارة:** قام معمل اختبارات ـ مجّهز بأجهزة محاكاة للقيادة على الطرق السريعة بدراسة العلاقة بين تكلفة التشغيل لكل ميل Y وسرعة السير (X_1) لنوعين من عجلات الشاحنات. ويوضح الجدول التالي المشاهدات (جميع البيانات مرمّزة) ويرغب أحد المهندسين في تقدير ما إذا كان انحدار التكلفة على السرعة نفسه لكل من النوعين أم لا. افترض أن تباينات الخطأ نفسها للنوعين وأن النموذج (10.6) هو النموذج المناسب.

النوع A

10	9	8	7	6	5	4	3	2	1	i
70	60	60	50	40	40	30	20	20	10	X_{i1}
25.8	24.1	22.4	20.9	16.5	19.0	14.9	14.2	12.5	9.8	Y_i

النوع B

10	9	8	7	6	5	4	3	2	1	i
70	60	60	50	40	40	30	20	20	10	X_{i1}
21.4	19.8	22.3	20.9	19.1	16.4	16.5	16.1	14.5	15.0	Y_i

أ ـ ارسم بيانيا في شكل واحد بيانات العينة لكل من المجتمعين مستخدما رموزا مختلفة لكل من العينتين. هل تبدو العلاقة بين السرعة والتكلفة نفسها لكل من نوعي الإطارات؟

ب ـ اختبر ما إذا كانت دوال الانحدار هي نفسها لكل من نوعي الإطارات أم لا. اضبط الخطأ من النوع الأول عند 0.05. اكتب البدائل، قاعدة القرار والنتيجة.

جـ ـ افترض أن السؤال المثير للاهتمام هو ببساطة ما إذا كان لخطي الانحدار ميلان متساويان. أجب على السؤال بتقدير الفرق بين الميلين مستخدما فترات ثقة بمعامل ثقة 0.95. ماذا تجد؟

د ـ احسب الرواسب وارسمها في مقابل $\hat{Y}$ وذلك بصورة منفصلة لكل نوع من الإطارات. هل تبدو فرضية تساوي تباينات الخطأ مقبولة هنا؟

هـ - مستخدما اختبار F. اختبر تساوي تبايني الخطأ لكل نوع من الإطارات مستخدما $\alpha = 0.01$ واكتب البدائل وقاعدة القرار والنتيجة. ماهي القيمة -P لهذا الاختبار؟

و - ارسم بيانيا في أشكال مختلفة الرواسب في مقابل X_1، X_2 و$X_2 X_1$ وقم أيضا بإعداد رسم احتمال طبيعي للرواسب. لخّص نتائج تحليلك لتلك الرسومات.

(١٠-١٦) بالإشارة إلى **كتلة العضلة** مسألة (٢-٢٥)، لدى اختصاصي التغذية حدس بأن انحدار كتلة العضلة على العمر يتبع علاقة خطية من قطعتين بميل يتغير عند عمر 60 وبدون انقطاع.

أ - اكتب نموذج الانحدار الذي يمكن تطبيقه إذا كان حدس اختصاصي التغذية سليما. ماهي دوال الاستجابة المقابلة حينما يكون العمر أقل أو يساوي 60 سنة، وحينما يكون العمر أكبر من 60؟.

ب - قم بتوفيق نموذج الانحدار الموصوف في (أ) واكتب دالة الانحدار المقدّرة.

جـ - اختبر الحاجة إلى انحدار خطي بقطعتين، استخدم $\alpha = 0.01$. اكتب البدائل، قاعدة القرار والنتيجة، ماهي القيمة -P لهذا الاختيار؟

(١٠-١٧) **معالجة الشحنات**. تستورد شركة عالمية للالكترونيات — على فترات — شحنات من القطع الكبيرة التي تُستخدم كمركّب في العديد من منتجاتها. ويختلف حجم الشحنة طبقا لجداول الإنتاج. ولمعالجة وتوزيع الشحنات على مراكز التجميع تم توزيعها كالآتي: أرسلت الشحنات من حجم أقل من أو يساوي 250000 قطعة إلى مستودع البضائع A، وأرسلت الشحنات الأكبر إلى مستودع البضائع B، حيث تتوافر فيه معدات متخصصة تحقق وفرا في معالجة الشحنات الكبيرة. تم جمع البيانات التالية لآخر 10 شحنات. وترمز X لحجم الشحنة (مقاسة بالألف جزء) وY للتكلفة المباشرة لمعالجة الشحنة في المستودع (بآلاف الدولارات).

10	9	8	7	6	5	4	3	2	1	i
125	400	290	325	180	175	200	150	350	225	X_1
7.97	15.00	13.30	13.75	10.13	10.03	10.98	8.93	14.13	11.95	Y_i

يراد توفيق نموذج انحدار خطي ذي قطعتين مع إمكانية انقطاع عند $X = 250$.

أ - حدّد نموذج الانحدار الذي يمكن استخدامه.

ب - قم بتوفيق ذلك النموذج وارسم دالة الاستجابة المقدَّرة والبيانات. هل هناك مايشير إلى تحقق وفر في معالجة الشحنات الكبيرة نسبيا بالمقارنة مع الشحنات الصغيرة ؟

جـ - اختبر ما إذا كان يمكن الاستغناء عن وجود ميلين مختلفين وانقطاع في النموذج أم لا. اضبط مستوى المعنوية عند 0.025. اكتب البدائل، قاعدة القرار والنتيجة.

د - بالنسبة للشحنات الصغيرة نسبيا. ماهو التقدير النقطي لتكلفة المعالجة المتوقعة لكل زيادة بمقدار ألف في حجم الشحنة؟ ماهو التقدير المقابل في حالة الشحنات الكبيرة نسبيا؟

(١٠-١٨) في تحليل السلاسل الزمنية: يُعرَّف المتغير X - الذي يمثل الزمن - عادة بحيث يأخذ القيم 1، 2، 3،..الخ. لفترات زمنية متتابعة، هل يمثّل هذا ترميزا محددا حينما تكون فترات الزمن هي بالفعل 1989، 1990،... الخ...؟

(١٠-١٩) رغب أحد المحللين في اعتبار عدد الأخوة الأكبر كمتغير مستقل في تحليل انحدار للعوامل المؤثرة على درجة النضج لطلبة الصف الثامن. ويتراوح عدد الأخوة الأكبر في مشاهدات العينة من 0 إلى 4. ناقش ما إذا كان يمكن تمثيل هذا المتغير كمتغير كمّي عادي أو بواسطة أربعة متغيرات مؤشرة 0 و1.

(١٠-٢٠) بالإشارة إلى نموذج انحدار (١٠-٢) لدراسة خطة التأمين المبتكرة، افترض أن X_0 أُسقطت لاستبعاد عدم الاستقلال الخطي في المصفوفة **X** وبذلك يصبح النموذج $Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$. ماهو معنى كل من معالم الانحدار β_1، β_2 وβ_3 هنا؟

تمارين

(١٠-٢١) بالإشارة إلى مثال دراسة معايرة الأداة في الفقرة (١٠-٤).

افترض أن هناك ثلاث أدوات (A، B وC) تم تطويرها بمواصفات متطابقة، وبحيث إن دوال الانحدار التي تربط قراءة المعايرة Y بالضغط الفعلي X_1 لكل أداة هي كثيرة حدود من الدرجة الثانية وافترض أن تباينات حد الخطأ متساوية وأن معاملات كثيرة الحدود قد تختلف من أداة إلى التي تليها. ليكن X_3 تعبيرا عن المتغير المؤشر الثاني بحيث يكون $X_3 = 1$ إذا كانت C الأداة و 0 فيما عدا ذلك.

أ - وسّع نموذج الانحدار (10.21) ليغطي هذه الحالة.

ب - اكتب البدائل وعرّف إحصاءة الاختبار واكتب قاعدة القرار لكل من الاختبارات التالية عندما يكون $\alpha = 0.01$:

(١) اختبر ما إذا كانت دوال الانحدار من المرتبة الثانية للأدوات الثلاث متطابقة، (٢) اختبر ما إذا كان لدوال الانحدار الثلاث الجزء المقطوع نفسه و(٣) اختبر ما إذا كانت كل من التأثيرات الخطية والتربيعية نفسها في دوال الانحدار الثلاث جميعها.

(١٠-٢٢) بالإشارة إلى المسألة (١٠-١٦) **كتلة العضلة** حدّد نموذج الانحدار حينما يتغير الميل عند عمر 40 سنة ومرة أخرى عند عمر 60 سنة بدون انقطاع.

(١٠-٢٣) في دراسة انحدار شملت ثلاثة أنواع من المصارف هي تجاريه، ادخار تعاوني، وادخار وتسليف، اعتبر النظام التالي للمتغيرات المؤشرة الخاصة بنوع المصرف:

نوع المصرف	X_2	X_3
تجاري	1	0
ادخار تعاوني	0	1
ادخار وتسليف	-1	-1

أ ـ طوّر نموذج انحدار خطّي من المرتبة الأولى لربط ربح أو خسارة العام الماضي Y بحجم المصرف (X_1) ونوع المصرف (X_3, X_2).

ب ـ اكتب دوال الاستجابة لأنواع المصارف الثلاثة.

جـ ـ فسّر معنى كل من الكميات التالية: (١) β_2، (٢) β_3 و(٣) $\beta_2 - \beta_3$.

(١٠-٢٤) بالإشارة إلى نموذج الانحدار (10.17) ومع استبعاد المتغير X_3

أ ـ اكتب المصفوفة $\mathbf{X'X}$ لهذه الحالة الخاصة بمتغير مستقل نوعي واحد. وذلك من أجل $i = 1, ..., n$ و n_1 من الشركات غير المحدودة.

ب ـ أوجد $\mathbf{b}$ باستخدام (7.21)

جـ ـ أوجد SEE و SSR باستخدام (7.30) و (7.31)

مشاريع

(١٠-٢٥) بالإشارة إلى مجموعة بيانات SMSA، يراد دراسة انحدار عدد الأطباء الممارسين Y على عدد السكان الكلي (X_1) الدخل الشخصي الإجمالي، (X_2) والمناطق الجغرافية.

أ ـ قم بتوفيق نموذج انحدار من المرتبة الأولى. ليكن $X_3 = 1$ إذا كانت المنطقة الجغرافية NE و 0 فيما عدا ذلك، و $X_4 = 1$ إذا كانت المنطقة NC و 0 فيما عدا ذلك، $X_5 = 1$ إذا كانت المنطقة S و 0 فيما عدا ذلك.

ب ـ اختبر ما إذا كان تأثير منطقة NE على عدد الأطباء الممارسين مختلفا عن تأثير منطقة NC وذلك بوضع فترة ثقة مناسبة بمعامل ثقة 0.90. فسّر التقدير بفترة الذي حصلت عليه.

جـ ـ اختبر ما إذا كانت التأثيرات الجغرافية موجودة، استخدم $\alpha = 0.10$. اكتب البدائل، قاعدة القرار والنتيجة. ماهي القيمة Pـ لهذا الاختبار؟

(١٠-٢٦) بالإشارة إلى مجموعة بيانات **SENIC**. يراد دراسة انحدار خطـورة الإصابـة Y على طـول فـترة الإقامـة (X_1) والعمـر (X_2)، ونسـبة أشـعة X الروتينيـة للصدر (X_3) والتبعية لمدرسة طبية (X_4):

أ ـ قم بتوفيق نموذج انحدار مـن المرتبـة الأولى، وليكـن $X_4 = 1$ اذا كـان المستشفى يتبع مدرسة طبية و 0 فيما عدا ذلك.

ب ـ قدِّر تأثير التبعية لمدرسـة طبيـة علـى خطـورة الإصابـة مسـتخدما 98 بالمائة فترة ثقة. فسّر معنى التقدير بفترة الذي حصلت عليه.

جـ ـ من المقترح أن تأثير التبعيـة لمدرسـة طبيـة على خطـورة الإصابـة قـد يتفاعل مع تأثيرات العمر ونسـبة أشـعة X الروتينيـة للصـدر. أضـف حدود التفاعل الملائمة لنموذج الانحدار. قم بتوفيـق نمـوذج الانحـدار المحسّن واختبر ما إذا كانت حدود التفـاعل مفيـدة، استخدم $\alpha = 0.10$. اكتب البدائل وقاعدة القرار والنتيجة.

(١٠-٢٧) بالإشارة إلى مجموعة بيانات **SENIC** يراد دراسـة انحـدار طـول الإقامـة Y علـى العمـر (X_1) ونسـبة الـزرع الروتيـني (X_2) ومتوسـط التعـداد اليومـي (X_3)، والتسهيلات والخدمات المتاحة (X_4) والمنطقة (X_5، X_6 وX_7).

أ ـ قم بتوفيق نموذج انحدار من المرتبة الأولى، ليكـن $X_5 = 1$ إذا كـانت المنطقة NE و 0 فيما عدا ذلك، $X_6 = 1$ إذا كـانت المنطقـة NC و 0 فيما عدا ذلك، $X_7 = 1$ إذا كانت المنقطة S و 0 فيما عدا ذلك.

ب ـ اختبر ما إذا كان من الممكـن إلغـاء نسبة الـزرع الروتيـني مـن النمـوذج، استخدم مستوى معنوية 0.05. اكتب البدائل وقاعدة القرار والنتيجة.

جـ ـ اختبر ما إذا كان تأثير طول الإقامة في المستشفيات الواقعة في منطقـة الغرب يختلف عنه في المستشفيات الواقعة في المناطق الثلاث الأخرى، وذلك بوضع فترة ثقة مناسبة لكـل مقارنـة بـين منطقتـين. استـخدم أسلوب بونفرّوني بمعامل ثقة عائلة 95 في المائة ، لخّص نتائجك.

تشخيصات وتدابير علاجية - II

نتابع في هذا الفصل عددا من التشخيصات المحسّنة للتحقق من صلاحية نموذج انحدار. وتتضمن هذه فيما تتضمن طرقا لكشف عدم صلاحية نموذج من حيث شموله لمتغير مستقل، المشاهدات القاصية، المشاهدات المؤثرة، والخطية المتعددة. وندرس أيضا تدابير علاجية متنوعة لمعالجة مثل هذه المشاكل، وتتضمن هذه التدابير انحدار الحافة الخاص بالخطية المتعددة والمربعات الدنيا المرجحة الخاصة بتباينات غير متساوية للخطأ.

(١١-١) صلاحية نموذج لمتغير مستقل - رسوم الانحدار الجزئي

ناقشنا في الفصلين الرابع والسابع كيف يمكن استخدام رسوم الرواسب في مقابل متغير مستقل في نموذج الانحدار للتحقق مما إذا كنا نحتاج في النموذج إلى تأثير منحن لذلك المتغير المستقل، ولفحص ما إذا كان تباين حدود الخطأ يتغير بطريقة نمطية مع المتغير المستقل. وقد وصفنا هناك أيضا رسم الرواسب في مقابل متغيرات مستقلة لم يشملها نموذج الانحدار بعد وذلك لتحديد ما إذا كان من المفيد إضافة واحد أو أكثر من هذه المتغيرات إلى النموذج.

ومحدودية رسوم الرواسب هذه هي أنها إذ تبين متى تكون العلاقة الخطية في متغير مستقل غير صالحة، فإنها لاتعرض طبيعة العلاقة في هذا المتغير المستقل التي ينبغي تمثيلها في نموذج الانحدار.

ورسوم الانحدار الجزئي هي رسوم رواسب محسّنة تبيّن العلاقة المناسبة لمتغير مستقل في نموذج الانحدار، وهي بالتالي تتمة قيّمة لرسوم الرواسب المعتادة. وصفة الجزئي في رسوم الانحدار الجزئي تأتي بمعنى أنها تعتبر الدور الهامشي الذي يلعبه متغير مستقل X_k . علما أن المتغيرات المستقلة الأخرى المعنية مشمولة من حينها في النموذج. وهكذا ينبغي استخدام رسوم الانحدار الجزئي بحذر، فهي تستخدم بصورة أولية لتقديم

معلومات عن التمثيل الدالي والأهمية الهامشية لمتغير مستقل نريد إضافته إلى نموذج الانحدار. وهذا التحذير مشابه لتحذير كنا ناقشناه في الفصل الثامن حول الاختبارات المتزامنة لكل معامل بمفرده من معاملات الانحدار، باعتبار أن هذه الاختبارات هي أيضا هامشية في طبيعتها.

وفي رسم الانحدار الجزئي، نحدر كلا من المتغير التابع Y والمتغير المستقل المعني X_k على المتغيرات المستقلة الأخرى في نموذج الانحدار ونحصل على الرواسب لكل منهما. وتعكس هذه الرواسب الجزء من كل متغير الذي لم يقترن خطيا بالمتغيرات المستقلة الأخرى، التي يشملها النموذج. ورسم مجموعتي الرواسب هاتين، إحداهما في مقابل الأخرى يكشف عن (١) طبيعة علاقة الانحدار للمتغير المستقل X_k موضع الدراسة من حيث إمكانية شموله في نموذج الانحدار و(٢) الأهمية الهامشية لهذا المتغير في تخفيض تشتت الراسب.

ولجعل هذه الأفكار أكثر تحديدا، دعنا ندرس نموذج انحدار متعدد بمتغيرين مستقلين X_1 و X_2 والتعميم إلى أكثر من متغيرين مستقلين هو تعميم مباشر. لنفرض أننا نهتم بطبيعة تأثير الانحدار الخاص بـ X_1 علما أن النموذج يتضمن X_2. نحدر Y على X_2 ونحصل على القيم التوفيقية والرواسب:

$$\hat{Y}_i(X_2)=b_0+b_2X_{i2} \quad (11.1a)$$

$$e_i(Y|X_2)=Y_i-\hat{Y}_i(X_2) \quad (11.1b)$$

وتشير الرموز هنا إلى المتغيرين التابع والمستقل في النموذج التوفيقي، ونحدر أيضا X_1 على X_2 لنجد:

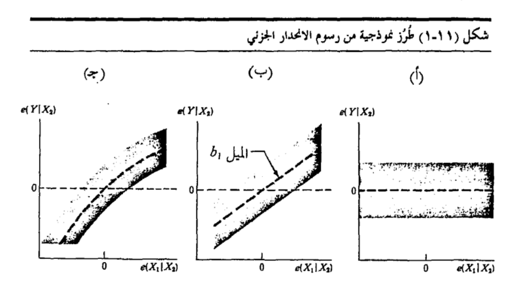
$$\hat{X}_{i1}(X_2)=b_0^*+b_2^*X_{i2} \quad (11.2\ a)$$

$$e_i(X_1|X_2)=X_{i1}-\hat{X}_{i1}(X_2) \quad (11.2\ b)$$

ويتألف رسم الانحدار الجزئي للمتغير المستقل X_1 من رسم الرواسب $e(Y \mid X_2)$ في مقابل رواسب X_1 وهي $e(X_1 | X_2)$.

ويتضمن الشكل (١١-١) عدة طُرُز نموذجية من رسوم الانحدار الجزئي، بما يتفق ومثالنا، حيث X_2 موجودة في النموذج وإضافة X_1 إلى النموذج هي موضع الدراسة. ويبين الشكل (١١-١)أ عصابة أفقية، تشير إلى أن X_1 لاينطوي على أية علاقة انحدار إضافية للتنبؤ بـ Y وراء تلك المحتواة في X_2، وبالتالي فلا فائدة تُرجى هنا من إضافة X_1 إلى نموذج الانحدار.

شكل (١١-١) طُرُز نموذجية من رسوم الانحدار الجزئي



ويبين الشكل (١١-١)ب عصابة خطية بميل غير الصفر. ويشير هذا إلى أن حدا خطيا في X_1 يمكن أن يشكل إضافة قيّمة إلى نموذج انحدار تضمن من حينه X_2. ويمكن تبيان أن ميل خط المربعات الدنيا المار من المبدأ في رسم الرواسب (١١-١)ب هو b_1 معامل الانحدار لِـ X_1 لو أن المتغير X_1 أضيف إلى نموذج الانحدار الذي تضمن من حينه X_2.

وتمثل الانحرافات الرأسية للنقاط المرسومة حول الخط الأفقي $e(Y \mid X_2) = 0$ في الشكل (١١-١)ب، رواسب Y عندما يكون X_2 في نموذج الانحدار. وعند تربيع وجمع هذه الانحرافات نحصل على مجموع مربعات الخطأ $SSE(X_2)$. ويمكن تبيان أن الانحرافات الرأسية للنقاط المرسومة مأخوذة حول الخط المار من المبدأ وميله b_1 هي الرواسب $e(Y|X_1, X_2)$ التي نجدها عندما يكون كل من X_1 و X_2 في نموذج الانحدار.

وبالتالي فإن مجموع مربعات هذه الانحرافات هو مجموع مربعات الخطأ $SSE(X_1, X_2)$. والفرق بين مجموعي المربعات هذين هو، وفقا لـ (8.1a) مجموع المربعات الإضافي $SSR(X_1|X_2)$. وهكذا فإنه إذا كان تبعثر النقاط حول الخط المار من المبدأ بميل يساوي b_1 أقل بكثير من تبعثرها حول الخط الأفقي، فإن ضم X_1 إلى نموذج الانحدار سيقدم تخفيضا إضافيا قيّما لمجموع مربعات الخطأ.

ويبين الشكل (١١-١)جـ عصابة منحنية مما يشير إلى أن إضافة X_1 إلى نموذج الانحدار يمكن أن تكون مفيدة، وينبغي لهذه الإضافة أن تنطوي على تأثير منحنٍ من النوع الذي يبينه النمط المعروض. ورسوم الانحدار الجزئي مفيدة أيضا في الكشف عن نقاط قاصية أو مشاهدات مؤثرة.

مثال ١.

يبين الجدول (١١-١) متوسط الدخل السنوي خلال السنتين الماضيتين X_1، درجة تجنّب المخاطرة X_2 ومقدار التأمين على الحياة Y. لعينة من 18 مديرا تتراوح أعمارهم بين 30 و 39 عاما. وقد قيس تجنب المخاطرة من خلال استبيان قياسي قُدّم لكل مدير: كلما كانت الدرجة أعلى كلما كان الحرص على تجنب المخاطرة أكبر. والدخل وتجنّب المخاطرة مرتبطان ارتباطا معتدلا هنا، فمعامل الارتباط $r_{12} = 0.245$.

ويعطي توفيق نموذج انحدار من المرتبة الأولى (الحسابات غير مبينة) مايلي:

$$\hat{Y} = -205.72 + 6.2880X_1 + 4.738X_2 \quad (11.3)$$

وقد رُسمت رواسب هذا النموذج التوفيقي في مقابل X_1 في الشكل (١١-٢)أ. ويبين رسم الرواسب بوضوح الحاجة إلى تأثير منحن لـ X_1 في نموذج الانحدار.

ولدراسة طبيعة هذا التأثير، سنستخدم رسم الانحدار الجزئي، ونحدر كلا من Y و X_1 على X_2. وعند القيام بذلك، نحصل (الحسابات غير مبينة) على:

$$\hat{Y}(X_2) = 50.70 + 15.54X_2 \quad (11.4a)$$

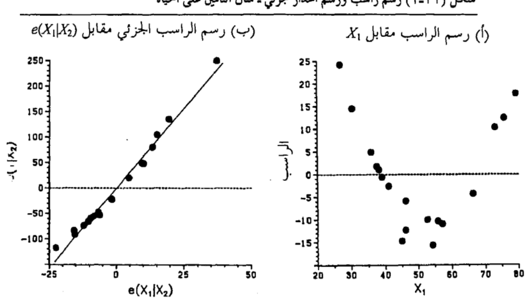
$$\hat{X}_1(X_2) = 40.779 + 1.718X_2 \quad (11.4b)$$

جدول (١١-١) بيانات مثال التأمين على الحياة (بآلاف الدولارات)

المدير i	متوسط الدخل السنوي بآلاف الدولارات X_{i1}	درجة تجنب المخاطرة X_{i2}	مقدار بوليصة التأمين (بآلاف الدولارات) Y_i
1	66.290	7	240
2	40.964	5	73
3	72.996	10	311
4	45.010	6	91
5	57.204	4	162
6	26.852	5	11
7	38.122	4	54
8	35.840	6	53
9	75.796	9	326
10	37.408	5	55
11	54.376	2	130
12	46.186	7	112
13	46.130	4	91
14	30.366	3	14
15	39.060	5	63
16	79.380	1	316
17	52.766	8	154
18	55.916	6	164

شكل (١١-٢) رسم راسب ورسم انحدار جزئي - مثال التأمين على الحياة

(أ) رسم الراسب مقابل X_1 (ب) رسم الراسب الجزئي مقابل $e(X_1|X_2)$



وقد رُسمت رواسب هذين النموذجين التوفيقيين الواحدة في مقابل الأخرى في رسم الانحدار الجزئي في الشكل (١١-٢)ب. ويتضمن هذا الرسم أيضا خط المربعات الدنيا عبر المبدأ بميل $b_1 = 6.2880$. ويبين رسم الانحدار الجزئي طبيعة العلاقة المنحنية بين Y و X_1 عندما كان X_2 من حينه في نموذج الانحدار، ونقصد أنها محدبة. وهذا واضح من الانحرافات الرأسية حول الخط المار من المبدأ بميل يساوي b_1 وهذه الانحرافات موجبة على اليسار، سالبة في الوسط، وموجبة ثانية على اليمين. وبصورة إجمالية، فإن التأثير المنحني متواضع ضمن مدى المتغيرات المستقلة.

ونلاحظ أيضا أن الانحرافات الرأسية حول العلاقة المنحنية عند توفيقها، ستكون أصغر بكثير من الانحرافات الرأسية حول الخط الأفقي، مما يشير إلى أن إضافة X_1 إلى نموذج الانحدار وفق علاقة منحنية ستخفض مجموع مربعات الخطأ تخفيضا كبيرا. وفي الحقيقة، فإن معامل التحديد الجزئي لتأثير خطّي لـ X_1 هو $r^2_{Y1|2} = 0.984$.

مثال ٢.

في مثال شحوم الجسم في الجدول (٨-١) (صفحة ٣٤٨)، نستعرض هنا انحدار شحوم الجسم Y على سماكة جلد العضلة ثلاثية الرؤوس X_1 ومحيط الفخذ X_2 فقط. ونحذف المتغير المستقل الثالث (X_3، محيط منتصف الذراع) وذلك كي نركّز مناقشة رسوم الانحدار الجزئي على عناصرها الأساسية، ونتذكر أن X_1 وX_2 مرتبطان ارتباطا عاليا $r_{12} = 0.92$. وقد حصلنا على دالة الانحدار التوفيقية في الجدول (٨-٢)جـ (صفحة ٣٤٩).

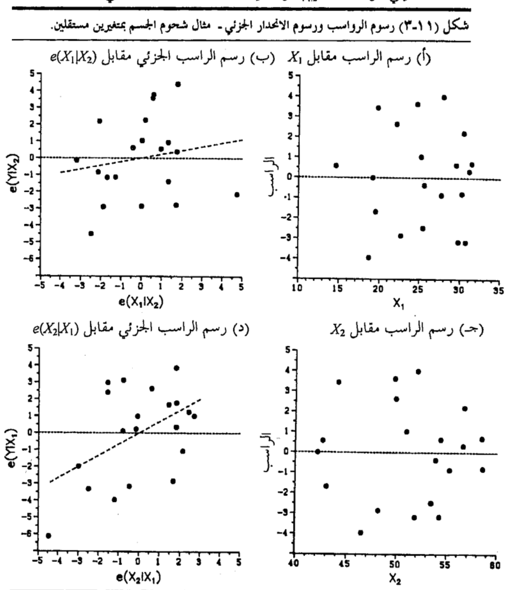
$$\hat{Y} = -19.174 + 0.2224 X_1 + 0.6594 X_2$$

ويتضمن الشكلان (١١-٣)أ و(١١-٣)جـ رسوم الرواسب في مقابل X_1 و X_2، على الترتيب، ولاتشير هذه الرسوم إلى أي نقص في توفيق الحدود الخطية في نموذج الانحدار أو إلى وجود تباينات غير متساوية لحدود الخطأ.

ويتضمن الشكلان (١١-٣)ب و(١١-٣)د رسوم الانحدار الجزئي لـ X_1 و X_2، على الترتيب، حيث كان المتغير المستقل الآخر من حينه في النموذج. ويبين الشكلان

أيضا الخط عبر المبدأ بميل يساوي معامل الانحدار للمتغير المستقل الآخر فيما لو أضيف إلى النموذج التوفيقي. والتبعثر في الشكل (١١ـ٣)ب يتبع الطراز المبين في الشكل (١١ـ١)أ، وحقيقة أن الفائدة الهامشية لـ X_1 تبدو طفيفة، عندما يكون X_2 في نموذج الانحدار، تتفق مع ما وجدناه سابقا في الفصل الثامن. وقد رأينا هناك أن معامل التحديد الجزئي هو $r^2_{Y1.2} = 0.031$ وأن الإحصاءة t^* الخاصة بـ b_1 هي 0.73 فقط.

شكل (١١ـ٣) رسوم الرواسب ورسوم الانحدار الجزئي ـ مثال شحوم الجسم بمتغيرين مستقلين.



ويتبع رسم الانحدار الجزئي لـ X_2 في الشكل (١١-٣)د، الطراز المبين في الشكل (١١-١)ب، مبينا تبعثرا خطيا بميل موجب، ونرى أيضا أن هناك، إلى حد ما، متغيرية حول الخط الذي ميله b_2 أقل من المتغيرية حول الخط الأفقي. وهذا يقترح: (١) يمكن أن يكون المتغير X_2 مفيدا في نموذج الانحدار حتى عندما يكون X_1 من حينه في النموذج. (٢) يبدو أن حدا خطيا في X_2 هو المناسب إذ ليس هناك ما يشير إلى وجود علاقة منحنية. وهكذا يدعم رسم الانحدار الجزئي لـ X_2 في الشكل (١١-٣)د رسم الراسب المعتاد في الشكل (١١-٣)جـ من حيث إنه يشير إلى الفائدة الجمّة لوجود محيط الفخذ X_2 في نموذج الانحدار عندما تكون سماكة جلد العضلة ثلاثية الرؤوس X_1 من حينها في النموذج، وتتسق هذه المعلومات مع كون الإحصاءة t^* لـ b_2، في الجدول (٨-٢)جـ، مساوية لِـ 2.26 وكون معامل التحديد الجزئي $r_{Y2.1}^2 = 0.232$ معتدلا.

تعليقات

١- يُظهر رسم الانحدار الجزئي، بصورة بيانية، طبيعة العلاقة الدالية التي ينبغي بموجبها إضافة متغير إلى نموذج الانحدار. وهو لايقدم عبارة تحليلية لهذه العلاقة. وتتوفر، في الغالب، تشكيلة من التحويلات أو حدود التأثير المنحني لتمثيل العلاقة التي يُظهرها الرسم. وسنحتاج إلى تقصّي هذه البدائل واستخدام المزيد من رسوم الرواسب لتحديد التحويل أو حدود التأثير المنحني الأفضل.

٢- عندما يتطلب الأمر عدة رسوم انحدار جزئية لمجموعة من المتغيرات المستقلة، فليس من الضروري توفيق نماذج انحدار جديدة كليا في كل مرة. وهناك طرق حسابية توفر في الحسابات المطلوبة، ومثل هذه الطرق مشروحة في كتب مدرسية مختصة مثل المرجع [11.1].

٣- يمكن الحصول على أية دالة انحدار متعدد توفيقية من متتالية من الانحدارات الجزئية التوفيقية. وللتوضيح، لنعتبر ثانية مثال التأمين على الحياة، حيث تعطي (11.4a) انحدارا توفيقيا لِـ Y على X_2، وتعطي (11.4b) انحدارا توفيقيا لـ X_1 على X_2. إذا حدرنا الآن الرواسب $Y - \hat{Y}(X_2)$ على الرواسب $X_1 - \hat{X}_1(X_2)$ مستخدمين الانحدار عبر المبدأ، فإننا نحصل (الحسابات غير مبينة) على:

$$\widehat{\left[Y-\hat{Y}(X_2)\right]}=6.2880\left[X_1-\hat{X}_1(X_2)\right] \qquad (11.5)$$

وبتعويض بسيط، مستخدمين (11.4a) و(11.4b) نجد

$$[Y - (50.70 + 15.54X_2)] = 6.2880[X_1 - (40.779 + 1.718X_2)]$$

أو:

$$\hat{Y}=-205.72+6.2880X_1+4.737X_2 \qquad (11.6)$$

حيث الحل من أجل Y هو القيمة التوفيقية $\hat{Y}$ عندما يشمل نموذج الانحدار X_1 و X_2. لاحظ أن النتيجة في (11.6) هي نفسها كما في توفيق نموذج الانحدار لـ X_1 و X_2 مباشرة، المعطى في (11.3) باستثناء فروق بسيطة تعود إلى آثار التدوير.

٤- ورسم الراسب الذي يتصل اتصالا وثيقا برسم الانحدار الجزئي، والمفيد بدوره في تحديد طبيعة العلاقة في متغير مستقل X_k ندرس إمكانية إضافته إلى نموذج انحدار، هو رسم الراسب الجزئي. ورسم الراسب هذا يأخذ كنقطة بداية الرواسب المعتادة $e_i=Y_i-\hat{Y}_i$ ويضاف إليها تأثير الانحدار لـ X_k وعلى وجه التحديد، نعرف الرواسب الجزئية، التي تهدف، إلى فحص تأثير المتغير المستقل X_k، ونرمز لها بـ $p_i(X_k)$، كما يلي:

$$p_i(X_k) = e_i + b_kX_{ik} \qquad (11.7)$$

وهكذا، للحصول على راسب جزئي، نضيف تأثير X_k كما يعكسه الحد b_kX_{ik} في النموذج التوفيقي، إلى الراسب. ويُشار إلى رسم هذه الرواسب الجزئية في مقابل X_k، كرسم راسب جزئي. ولمزيد من التفاصيل حول رسوم الراسب الجزئي نعيد القارئ إلى المرجع [11.2].

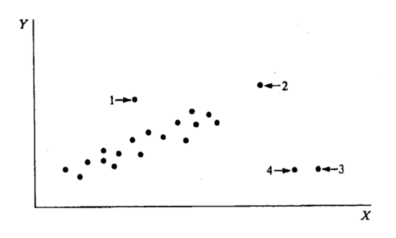
(١١-٢) تحديد مشاهدات قاصية في X - مصفوفة القبعة وقيم العزم.

حالات قاصية

كثيرا ماتحتوي مجموعة البيانات، في تطبيقات تحليل الانحدار، على بعض الحالات القاصية أو المتطرفة؛ أي أن المشاهدات الخاصة بهذه الحالات تكون منفصلة بوضوح عن بقية البيان الإحصائي. وقد تنطوي هذه الحالات القاصية على رواسب كبيرة

ويكون لها، في الغالب، تأثيرات دراماتيكية على دالة انحدار المربعات الدنيا التوفيقية. ومن المهم التالي دراسة الحالات القاصية بعناية وتقرير ما إذا كان ينبغي الاحتفاظ بها أو إلغاؤها، وفي حالة الاحتفاظ بها، تقرير ما إذا كان ينبغي تخفيض نفوذها في عملية التوفيق، و/أو إعادة النظر في نموذج الانحدار.

شكل (١١-٤) رسم انتشار لانحدار بمتغير مستقل واحد يوضح حالات قاصية



وقد تكون مشاهدة قاصية أو متطرفة بالنسبة للقيمة Y أو بالنسبة لقيمة (أو قيم) X، أو بالنسبة لكليهما. وهذا موضّح في الشكل (١١-٤) لحالة انحدار بمتغير مستقل واحد. ففي رسم الانتشار في الشكل (١١-٤) نجد أن المشاهدة 1 قاصية بالنسبة لقيمة Y. ونلاحظ أن هذه النقطة تقع بعيدا خارج مخطط الانتشار، مع أن قيمة X الخاصة بها قريبة من منتصف مدى المشاهدات بالنسبة للمتغير المستقل. والمشاهدات 2 ، 3 و4، هي مشاهدات قاصية بالنسبة للقيم X باعتبار أن القيم X الخاصة بها أكبر بكثير من تلك الخاصة بالمشاهدات الأخرى؛ والمشاهدتان 3 و4 قاصيتان أيضا بالنسبة لقيم Y.

وليس لجميع المشاهدات القاصية نفوذ قوي على دالة الانحدار التوفيقية. فقد لاتكون المشاهدة 1 في الشكل (١١-٤) ذات نفوذ كبير نظرا لوجود عدد من المشاهدات الأخرى بقيم مشابهة لـ X، مما سيحفظ دالة الانحدار التوفيقية من الانزلاق بعيدا كنتيجة للمشاهدة القاصية. وكذلك الأمر بالنسبة للمشاهدة 2، فقد لايكون لها

نفوذ قوي لأن قيمة Y فيها منسّقة مع علاقة الانحدار التي تُظهرها المشاهدات غير القاصية. وعلى الوجه الآخر، فمن المحتمل أن يكون للمشاهدتين 3 و4 نفوذ قوي من حيث تأثيرهما على دالة الانحدار التوفيقية، إنهما قاصيتان فيما يتعلق بقيمتي X فيهما، وقيمتا Y فيهما غير متسقتين مع علاقة الانحدار للمشاهدات الأخرى.

وإحدى الخطى الأساسية في أي تحليل انحدار هي تحديد ما إذا كان نموذج الانحدار المدروس خاضعا لسطوة مشاهدة واحدة أو قلة من المشاهدات في مجموعة البيانات. وفي انحدار بمتغير مستقل واحد أو متغيرين، يكون من السهل نسبيا التعرف على مشاهدات قاصية في قيم X أو في قيم Y بوسائل مثل الرسوم الصندوقية، رسوم الجذع والورقة، رسوم الانتشار، ورسوم الراسب، ودراسة ما إذا كان لها نفوذ له تأثيره على دالة الانحدار التوفيقية. إلا أنه عندما يشمل نموذج الانحدار أكثر من متغيرين مستقلين، يصبح التعرف على مشاهدة قاصية بالوسائل البيانية البسيطة أمرا صعبا، ذلك لأن تفحص متغير بمفرده أو متغيرين لايساعد بالضرورة على تحديد القاصيات بالنسبة لنموذج انحدار متعدد المتغيرات. وبعض القاصيات في متغير واحد قد لاتكون متطرفة في نموذج انحدار متعدد، وعلى العكس، قد لا نتمكن من اكتشاف قاصيات في عدد من المتغيرات عند تحليل يتطرق لمتغير واحد أو لمتغيرين منها.

ونناقش الآن استخدام مصفوفة القبعة المعرفة في (7.25a) للتعرف على مشاهدات قاصية في عدة متغيرات X. وفي الفقرة التالية نتابع بعض التدابير المحسّنة للتعرف على مشاهدات قاصية في قيم Y.

استخدام مصفوفة القبعة H للتعرف على مشاهدات قاصية في قيم X

واجهنا مصفوفة القبعة في الفصلين السادس والسابع:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{11.8}$$

وأشرنا في (7.25) إلى أنه يمكن التعبير عن القيم التوفيقية كتراكيب خطية في المشاهدات Y_i وذلك من خلال مصفوفة القبعة:

$$\hat{Y} = \mathbf{HY} \tag{11.9}$$

وبصورة مشابهة أشرنا في (7.27) إلى أنه يمكن التعبير عن الرواسب كتراكيب خطية في المشاهدات Y_i باستخدام مصفوفة القبعة:

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \qquad (11.10)$$

وفضلا عن ذلك، أشرنا في (7.27) إلى أن تباينات وتغايرات الرواسب تنطوي على مصفوفة القبعة:

$$\sigma^2\{\mathbf{e}\} = \sigma^2(\mathbf{I} - \mathbf{H}) \qquad (11.11)$$

وهكذا فإن تباين الراسب e_i، ونرمز له بـ $\sigma^2\{e_i\}$، هو:

$$\sigma^2\{e_i\} = \sigma^2(1 - h_{ii}) \qquad (11.12)$$

حيث h_{ii} هو العنصر من القطر الرئيس لمصفوفة القبعة.

ويمكن الحصول على العنصر القطري h_{ii} لمصفوفة القبعة مباشرة من:

$$h_{ii} = \mathbf{X}_i' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i \qquad (11.13)$$

حيث تقابل $\mathbf{X}_i$ هنا المتجه $\mathbf{X}_h$ في (7.48) باستثناء أن $\mathbf{X}_i$ تخص المشاهدة i :

$$\underset{p\times 1}{\mathbf{X}_i} = \begin{bmatrix} 1 \\ X_{i,1} \\ \cdot \\ \cdot \\ \cdot \\ X_{i,p-1} \end{bmatrix} \qquad (11.13a)$$

لاحظ أن $\mathbf{X}_i'$ هي ببساطة الصف i من المصفوفة $\mathbf{X}$ المتعلق بالمشاهدة i .

وللعناصر القطرية h_{ii} في مصفوفة القبعة بعض الخواص المفيدة، ونذكر، على وجه الخصوص، خاصيتي أن قيمها تقع دائما بين الصفر والواحد وأن مجموعها يساوي p :

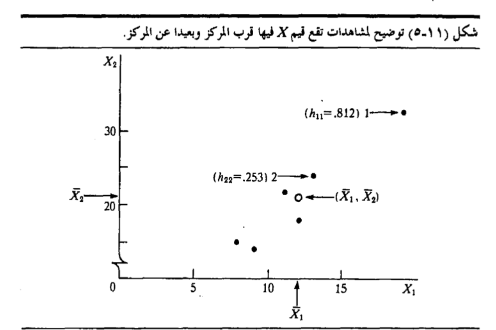
$$0 \le h_{ii} \le 1 \qquad \sum_{i=1}^{n} h_{ii} = p \qquad (11.14)$$

حيث p عدد معالم الانحدار في دالة الانحدار بما في ذلك حد الجزء المقطوع.

والعنصر القطري h_{ii} في مصفوفة القبعة هو مؤشر مفيد لما إذا كانت المشاهدة قاصية أم لا بالنسبة للقيم X، وذلك في دراسة متعددة المتغيرات. ويدعى العنصر القطري h_{ii} عزم المشاهدة i (بدلالة القيم X). وهو يشير إلى ما إذا كانت القيم X للمشاهدة i قاصية أم لا، ذلك لأنه يمكن تبيان أن h_{ii} يقيس المسافة بين قيم X

للمشاهدة i ومتوسطات القيم X للمشاهدات الـ n جميعها. وهكذا يشير كبر قيمة العزم h_{ii} إلى أن المشاهدة i بعيدة عن مركز المشاهدات X جميعا. ويوضح الشكل (١١ـ٥) هذا الأمر في حالة متغيرين مستقلين. فالمشاهدة 1 بعيدة عن المركز $(\overline{X}_1, \overline{X}_2)$ ولها قيمة عزم كبيرة $h_{ii} = 0.812$ بينما المشاهدة 2 قريبة من المركز ولها قيمة عزم صغيرة $h_{22} = 0.253$.

شكل (١١ـ٥) توضيح لمشاهدات تقع قيم X فيها قرب المركز وبعيدا عن المركز.



وإذا كانت المشاهدة i قاصية من حيث قيم X فيها وبالتالي لها قيمة عزم h_{ii} كبيرة. فإنها تُبدي عزما كبيرا في تحديد القيمة التوفيقية $\hat{Y}_i$ والأمر كذلك للأسباب التالية:

١ـ القيمة التوفيقية $\hat{Y}_i$ هي تركيب خطي في قيم Y الملحوظة، كما هو مبين في (11.9)، و h_{ii} هو وزن المشاهدة Y_i في تحديد هذه القيمة التوفيقية. وهكذا فإنه كلما كان h_{ii} أكبر كلما كانت Y_i أكثر أهمية في تحديد $\hat{Y}_i$. ولنذكر أن h_{ii} هي دالة في القيم X فقط، وبالتالي يقيس h_{ii} دور القيم X في تحديد مدى أهمية Y_i من حيث تأثيرها على القيمة التوفيقية $\hat{Y}_i$.

٢- كلما كان h_{ii} أكبر كلما كان تباين الراسب e_i أصغر، كما يمكن من (11.12) وبالتالي، كلما كان h_{ii} أكبر كلما مالت القيمة التوفيقية $\hat{Y}_i$ إلى أقرب للقيمة الملحوظة Y_i. وفي الحالة الحدية حيث $h_{ii}=1$، يكون $\sigma^2\{e_i\}=0$، وبال القيمة التوفيقية $\hat{Y}_i$ على أن تكون مساوية للقيمة الملحوظة Y_i، وبما أن المشاهدا العزم تميل إلى أن يكون راسبها أصغر، فقد لايكون من الممكن الكشف عنها إلى فحص الرواسب فقط.

وعادة تعتبر قيمة العزم h_{ii} كبيرة إذا تجاوزت ضعف متوسط قيم العزوم بـ $\bar{h}$، وهي تساوي وفقا للعلاقة (11.14):

$$\bar{h}=\frac{\sum_{i=1}^{n} h_{ii}}{n}=\frac{p}{n} \qquad (11.15)$$

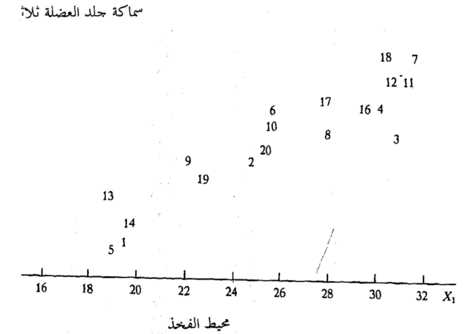
وبالتالي فإن قيم العزم الأكبر من $2p/n$ تُعتبر، وفقا لهذه القاعدة، مؤشر مشاهدات قاصية من حيث قيم X لهذه المشاهدات، والمرشد المقترح الآخر هـ قيم لـ h_{ii} تتجاوز 0.5 قيم عزم عالية جدا، بينما تعتبر القيم الواقعة بين 0.2 و5 عزم معتدلة. والبيّنة الإضافية لمشاهدة قاصية هي وجود ثغرة بين قيم العز المشاهدات وقيمة (أو قيم) عزم كبيرة بصورة غير عادية.

مثال

نستمر في مثال شحوم الجسم المبين في الجدول (٨-١)، ومن جديد فقط المتغيرين المستقلين: سماكة جلد العضلة ثلاثية الرؤوس X_1، ومحيط الف بحيث يمكن مقارنة نتائج استخدام مصفوفة القبعة بنتائج الرسوم البيانية ويتضمن الشكل (١١-٦) رسم انتشار لـ X_2 في مقابل X_1، حيث حُددت المشاهدات بالأرقام التي تشير إلى ترتيب الحصول عليها. ونلاحظ من الشكل (١١ المشاهدتين 15 و3 تبدوان قاصيتين بالنسبة إلى سير القيم X. والمشاهدة 15 قا X_1 وتشكل النهاية الدنيا في مدى X_2، بينما تبدو المشاهدة 3 قاصية في ا الخطية المتعددة، مع أنها ليست قاصية في أي من المتغيرين المستقلين بمفرده، المشاهدتان 1 و 5 متطرفتين إلى حد ما.

شكل (١١-٦) رسم انتشار لمحيط الفخذ في مقابل سماكة جلد العضلة ثلاثية الرؤوس في مثا

الجسم بمتغيرين مستقلين



وتؤيد حسابات مصفوفة القبعة (11.8) هذه الانطباعات إذ العمود(٢) من الجدول (١١-٢) قيم العزم لمثال شحوم الجسم ونلاحظ قيمتين للعزم هما $h_{33} = 0.372$ و $h_{15,15} = 0.333$. وكلاهما تتجاوز قاعد متوسط قيم العزم، $2p / n = 2(3)/20 = 0.30$، وتفصلهما ثغرة كبيرة العزم التالية لهما في الكبر، $h_{55} = 0.248$ و$h_{11} = 0.201$. وسنحتاج، وقا المشاهدتين 3 و 15 كمشاهدتين قاصيتين في قيم X، إلى التعرف على مد هاتين المشاهدتين في توفيق دالة الانحدار. وسنتعرض لهذا السؤال بعد أن نتا؛ تحديد المشاهدات القاصية في قيم Y.

جدول (١١-٢) الرواسب، العناصر القطرية لمصفوفة القبّعة، ورواسب الحذف المعيّرة تقديرا ـ مثال شحوم الجسم بمتغيرين مستقلين.

i	(١) e_i	(٢) h_{ii}	(٣) $\hat{d}_i$
1	-1.683	.201	-.730
2	3.643	.059	1.534
3	-3.176	.372	-1.656
4	-3.158	.111	-1.348
5	.000	.248	.000
6	-.361	.129	-.148
7	.716	.156	.298
8	4.015	.096	1.760
9	2.655	.115	1.117
10	-2.475	.110	-1.034
11	.336	.120	.137
12	2.226	.109	.923
13	-3.947	.178	-1.825
14	3.447	.148	1.524
15	.571	.333	.267
16	.642	.095	.258
17	-.851	.106	.344
18	-.783	.197	.335
19	-2.857	.067	-1.176
20	1.040	.050	.409

(١١-٣) تحديد مشاهدات قاصية في Y – رواسب الحذف المعيّرة تقديرا

درسنا في فصول سابقة الكشف عن مشاهدات قاصية أو متطرفة في Y استنادا إلى فحص الرواسب. وقد استخدمنا هناك إما الراسب e_i:

$$e_i = Y_i - \hat{Y}_i \qquad (11.16)$$

أو الرواسب المعيَّرة:

$$\frac{e_i}{\sqrt{MSE}} \qquad (11.17)$$

ونقدم الآن تحسينين يجعلان تحليل الرواسب أكثر فعالية في الكشف عن مشاهدات قاصية في Y.

رواسب معيّرة تقديرا بصورة داخلية

عندما يكون للرواسب e_i تباينات $\sigma^2\{e_i\}$ مختلفة اختلافا شديدا، حيث $\sigma^2\{e_i\}$ معطاة في (11.12)، فمن المستحسن، كي نأخذ في الحسبان الفروق في أخطاء المعاينة لكل منها، أن نعتبر مقدار e_i منسوبا إلى $\sigma\{e_i\}$. وقد رأينا في (7.28) أن المقدّر غير المنحاز لذلك التباين هو:

$$s^2\{e_i\} = MSE(1 - h_{ii}) \tag{11.18}$$

وتدعى نسبة e_i إلى $s\{e_i\}$ الراسب المعيَّر تقديرا بصورة داخلية وسنرمز له بـ e_i^*:

$$e_i^* = \frac{e_i}{s\{e_i\}} \tag{11.19}$$

وبينما سيكون للرواسب e_i تباينات معاينة مختلفة اختلافا شديدا إذا كانت قيم العزم h_{ii} مختلفة بصورة ملحوظة، فإن للرواسب المعيرة تقديرا بصورة داخلية تباينا ثابتا (عندما يكون النموذج مناسبا).

رواسب الحذف

والتحسين الثاني هو أن نقيس الراسب $e_i = Y_i - \hat{Y}_i$ عندما يُبنى الانحدار التوفيقي على المشاهدات بعد أن نستثني منها المشاهدة i. وسبب هذا التحسين هو أنه إذا كانت Y_i قاصية جدا فقد يؤثر هذا في دالة انحدار المربعات الدنيا التوفيقية المستندة إلى جميع المشاهدات ليجعلها أقرب إلى Y_i، مما يُنتج قيمة توفيقية $\hat{Y}_i$ قريبة من Y_i. وفي تلك الحالة، سيكون الراسب e_i صغيرا وسوف لايكشف عن كون Y_i قاصية. وعلى الوجه الآخر، إذا حُذفت المشاهدة i قبل توفيق دالة الانحدار، فإن قيمة المربعات الدنيا التوفيقية $\hat{Y}_i$ لاتتأثر بالمشاهدة القاصية Y_i، وسيميل الراسب عندئذ إلى أن يكون أكبر وبالتالي تزداد إمكانية كشفه للمشاهدة القاصية في قيمة Y.

والطريقة عندئذ هي أن نحذف المشاهدة، ثم نقوم بتوفيق دالة الانحدار للمشاهدات الـ $n - 1$ الباقية، ونقارن التقدير النقطي للقيمة المتوقعة عندما تكون مستويات X هي تلك الخاصة بالمشاهدة i المحذوفة، وسنرمز لها بـ $\hat{Y}_{i(i)}$، مع القيمة الملحوظة فعلا Y_i. ويذكّرنا الرمز $\hat{Y}_{i(i)}$ بأن المشاهدة i قد أُلغيت عند توفيق دالة الانحدار. ويدعى الراسب:

$$d_i = Y_i - \hat{Y}_{i(i)} \qquad (11.20)$$

راسب الحذف ونرمز له بـ d_i والعبارة المكافئة جبريا لـ d_i والتي لاتستدعي إعادة حساب دالة الانحدار التوفيقية بعد حذف المشاهدة i، هي:

$$d_i = \frac{e_i}{1 - h_{ii}} \qquad (11.20a)$$

حيث e_i الراسب المعتاد للمشاهدة i و h_{ii} هي قيمة العزم (11.13) لهذه المشاهدة ونلاحظ أنه كلما كانت قيمة العزم h_{ii} أكبر كلما كان راسب الحذف أكبر بالمقارنة مع الراسب المعتاد.

وهكذا فإن رواسب الحذف ستحدد أحيانا المشاهدات القاصية في قيم Y حيث تفشل الرواسب العادية في القيام بذلك ؛ وفي أحيان أخرى تقود رواسب الحذف إلى التحديدات نفسها التي تقود إليها الرواسب العادية.

ونلاحظ أن راسب الحذف يقابل خطأ التنبؤ في بسط العلاقة (3.34) عند التنبؤ بمشاهدة جديدة مستخدمين دالة الانحدار التوفيقية للمشاهدات السابقة، باستثناء أنه في (3.34) نجد الفرق $\hat{Y}_{i(i)} - Y_i$ والرموز تختلف عما هي عليه هنا. وهكذا نجد من (7.58a) أن التباين المقدّر لـ d_i هو:

$$s^2\{d_i\} = MSE_{(i)}(1 + \mathbf{X}_i'(\mathbf{X}_{(i)}'\mathbf{X}_{(i)})^{-1}\mathbf{X}_i) \qquad (11.21)$$

حيث $\mathbf{X}_i$ هو متجه المشاهدات X المذكور في (11.13) للمشاهدة i، و$MSE_{(i)}$ هو متوسط مربعات الخطأ في توفيق دالة الانحدار بعد الغاء المشاهدة i، و$\mathbf{X}_{(i)}$ هو المصفوفة $\mathbf{X}$ بعد حذف المشاهدة i. والعبارة المكافئة جبريا لِـ $s^2\{d_i\}$ هي:

$$s^2\{d_i\} = \frac{MSE_{(i)}}{1 - h_{ii}} \qquad (11.21a)$$

وينتج من (7.58) أن:

$$\frac{d_i}{s\{d_i\}} \sim t(n - p - 1) \qquad (11.22)$$

تذكّر أن $n - 1$ من المشاهدات قد استُخدمت هنا للتنبؤ بالمشاهدة i، وبالتالي فإن درجات الحرية هي $(n - 1) - p = n - p - 1$.

رواسب الحذف المعيّرة تقديرا

بضم التحسينين السابقين نحصل من أجل تشخيص المشاهدات القاصية أو المتطرفة في قيم Y على راسب الحذف d_i المعطى في (11.20) بعد معايرته معايرة تقديرية بقسمته على الانحراف المعياري المقدَّر المعطى في (11.21). وهكذا يكون راسب الحذف المعيّر تقديرا، وسنرمز له بـ d_i^* كما يلي:

$$d_i^* = \frac{d_i}{s\{d_i\}} \tag{11.23}$$

ونستنتج من (11.20a) و(11.21a) أن العبارة الجبرية المكافئة لـ d_i^* هي:

$$d_i^* = \frac{e_i}{\sqrt{MSE_{(i)}(1-h_{ii})}} \tag{11.23a}$$

ويُدعى الراسب المعيّر تقديرا في (11.23) أحيانا، الراسب المعيّر تقديرا بصورة خارجية ونعلم من (11.23) أن كل راسب حذف معيّر تقديرا d_i^* يتبع التوزيع t بـ $n - p - 1$ درجة من الحرية. وعلى أي حال، فإن المقادير d_i^* ليست مستقلة.

ومن حسن الحظ، يمكن حساب رواسب الحذف المعيّرة تقديرا d_i^* المعطاة في (11.23) دون الاضطرار إلى توفيق دوال انحدار جديدة في كل مرة نلغي فيها مشاهدة مختلفة. إذ توجد علاقة بسيطة بين MSE و $MSE_{(i)}$ هي:

$$(n-p)MSE = (n-p-1)MSE_{(i)} + \frac{e_i^2}{1-h_{ii}} \tag{11.24}$$

واستخدام هذه العلاقة في (11.23a) ينتج العبارة المكافئة التالية لـ d_i^*:

$$d_i^* = e_i\left[\frac{n-p-1}{SSE(1-h_{ii})-e_i^2}\right]^{1/2} \tag{11.25}$$

وهكذا يمكن حساب رواسب الحذف المعيّرة تقديرا d_i^* من الرواسب e_i، ومجموع مربعات الخطأ SSE، وقيم العزم h_{ii}، جميعها محسوبة من الانحدار التوفيقي القائم على n من المشاهدات.

ولتحديد المشاهدات القاصية في قيم Y، ننظر في مقدار القيم المطلقة لرواسب الحذف المعيرة تقديرا ثم نستخدم التوزيع t المناسب للتعرف على مدى وقوع مثل هذه القيم القاصية بعيدا في أحد الذيلين.

مثال. نوضّح حساب رواسب الحذف المعيّرة تقديرا بأخذ المشاهدة الأولى من مثال شحوم الجسم في الجدول (١١-٢) المتضمن للمتغيرين المستقلين X_1 و X_2. وقيم X لهذه المشاهدة، كما يعطيها الجدول (٨-١)، هي $X_{11} = 19.5$ و $X_{12} = 43.1$ وباستخدام دالة الانحدار التوفيقية في الجدول (٨-٢)جـ، نجد:

$$\hat{Y}_1 = -19.174 + 0.2224(19.5) + 0.6594(43.1) = 13.583$$

وبما أن $Y_1 = 11.9$ فالراسب لهذه المشاهدة هو $e_1 = 11.9 - 13.583 = -1.683$ ونعلم أيضا من الجدول (٨-٢)جـ. أن $SSE = 109.95$. ومن الجدول (١١-٢) أن $h_{11} = 0.201$.
وهكذا نجد من (11.25) أن:

$$d_1^* = -1.683\left[\frac{20-3-1}{109.95(1-.201)-(-1.683)^2}\right]^{1/2} = -.730$$

ورواسب الحذف المعيرة تقديرا لكل من المشاهدات العشرين مبينة في العمود الثالث من الجدول (١١-٢).

ونلاحظ أن رواسب الحذف المعيّرة تقديرا أو الأكبر في قيمتها المطلقة تعود للمشاهدات 3، 8 و13. وإذا اعتبرنا الذيلين على الجانبين، بمساحة 0.05 لكل منهما، متطرفين، فسنحتاج لمقارنة قيم رواسب الحذف المعيّرة تقديرا، بقيمة التوزيع t بـ $n - p - 1 = 16$ درجة من الحرية، أي المقارنة بـ $t(.95;16) = 1.746$ على وجه التحديد. وبالمصادفة فإن اعتبار الرواسب e_i هنا (وهي مبينة في العمود الأول من الجدول (١١-٢)) تكشف أيضا عن أن المشاهدتين 8 و13 هما المشاهدتان الأكثر قصوّا.

(١١-٤) تحديد المشاهدات المؤثّره - تدابير *DFFITS*، *DFBETAS* ومسافة كوك *COOK*

بعد تحديد المشاهدات القاصية بالنسبة لقيمها في X و/ أو قيمها في Y، تكون الخطوة التالية: هي التعرُّف على ما إذا كانت هذه المشاهدات القاصية مؤثّرة أم لا. وسنعتبر المشاهدة مؤثرة إذا كان استثناؤها يسبب تغيرات رئيسة في دالة الانحدار التوفيقية. وكما لاحظنا في الشكل (١١-٤)، لا حاجة لأن تكون جميع المشاهدات القاصية مؤثرة. وعلى سبيل المثال، قد لاتؤثر المشاهدة 1 في الشكل (١١-٤) في دالة الانحدار التوفيقية تأثيرا ذا بال.

وسنبدأ هنا بثلاثة مقاييس للتأثير، وهي مقاييس مستخدمة على نطاق واسع في التطبيق العملي، ويقوم كل منها على حذف مشاهدة واحدة لقياس تأثيرها.

التأثير على القيم التوفيقية ـ *DFFITS*

إحدى المقاييس المفيدة لتأثير المشاهدة على القيمة التوفيقية $\hat{Y}_i$ معطى بالعلاقة:

$$(DFFITS)_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}} \qquad (11.26)$$

ويرمز الحرفان DF للفرق بين القيمة التوفيقية $\hat{Y}_i$ للمشاهدة i عند استخدام جميع المشاهدات الـ n في توفيق دالة الانحدار، وبين قيمة التنبؤ $\hat{Y}_{i(i)}$ للمشاهدة التي نحصل عليها عند إلغاء المشاهدة i في عملية توفيق دالة الانحدار. وينطوي مقام (11.26) على متوسط مربعات الخطأ، عندما نلغي المشاهدة i في عملية توفيق دالة الانحدار، وعلى قيمة العزم h_{ii} المعرفة في (11.13). ويقدم المقام نوعا من المعايرة بحيث إن القيمة $(DFFITS)_i$، للنقطة i من نقاط البيان الإحصائي، تمثل بصورة تقريبية عدد الانحرافات المعيارية المقدَّرة التي تتغيرها القيمة التوفيقية $\hat{Y}_i$ عند إزاحة المشاهدة i من مجموعة البيانات.

ويمكن تبيان أنه يمكن حساب القيم ***DFFITS*** مستخدمين فقط النتائج المتوافرة من توفيق مجموعة البيانات بكاملها، وذلك وفقا للعلاقة التالية:

$$(DFFITS)_i = e_i \left[\frac{n-p-1}{SSE(1-h_{ii}) - e_i^2}\right]^{1/2} \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} = d_i^* \left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} \qquad (11.26a)$$

ونلاحظ من العبارة الأخيرة أن القيم ***DFFITS*** هي رواسب الحذف المعيّرة تقديرا بعد زيادتها أو إنقاصها من خلال عامل هو، في واقع الأمر، دالة في قيم العزم. وإذا كانت المشاهدة i قاصية في قيم X ولها قيمة عزم مرتفعة فإن هـذا العامل سيكون أكبر من الواحد وستنحو $(DFFITS)_i$ إلى أن تكون كبيرة بالقيمة المطلقة.

وكدليل لتحديد المشاهدات المؤثّرة، نقترح اعتبار المشاهدة مؤثرة إذا تجاوزت القيمة المطلقة لِـ ***DFFITS*** الواحد، من أجل مجموعات البيانات من حجم صغير إلى متوسط، وإذا تجاوزت $2\sqrt{p/n}$ في مجموعات كبيرة من البيانات.

مثال. في العمود الأول من الجدول (١١-٣) قائمة بقيم الـ ***DFFITS*** في مثال شحوم الجسم بمتغيرين مستقلين. ولتوضيح الحسابات، لنعتبر قيمة الـ ***DFFITS*** للمشاهدة 1. فقد وجدنا سابقا أن راسب الحذف المعيّر تقديرا لهذه المشاهدة هو $d_i^* = -0.730$ وأن قيمة العزم هي $h_{11} = 0.201$ وبالتالي نحصل، باستخدام (11.26a) على:

$$(DFFITS)_1 = -.730\left(\frac{.201}{1-.201}\right)^{1/2} = -0.366$$

جدول (١١-٣) ***DFFITS***، ***DFBETAS*** ومسافات كوك ـ مثال شحوم الجسم بمتغيرين مستقلين.

	(١)	(٢)	(٣)	(٤)	(٥)
		DFBETAS			
i	$(DFFITS)_i$	b_0	b_1	b_2	D_i
1	-.336	-.305	-.132	.232	.046
2	.384	.173	.115	-.143	.046
3	-1.273	-.847	-1.183	1.067	.490
4	-.476	-.102	-.294	.196	.072
5	.000	.000	.000	.000	.000
6	-.057	.040	.040	-.044	.001
7	.128	-.078	-.016	.054	.006
8	.575	.261	.391	-.333	.098
9	.402	-.151	-.295	.247	.053
10	-.364	.238	.245	-.269	.044
11	.051	-.009	.017	-.003	.001
12	.323	-.131	.023	.070	.035
13	-.851	.119	.592	-.390	.212
14	.636	.452	.113	-.298	.125
15	.189	-.003	-.125	.069	.013
16	.084	.009	.043	-.025	.002
17	-.118	.080	.055	-.076	.005
18	-.166	.132	.075	-.116	.010
19	-.315	-.130	-.004	.064	.032
20	.094	.010	.002	-.003	.003

وقيمة ***DFFITS*** الوحيدة، في الجدول (١١-٣)، التي تتجاوز الدليل لمجموعة بيانات من الحجم المتوسط هي تلك الخاصة بالمشاهدة 3، حيث $|(DFFITS)_3|=1.273$ وهذه القيمة هي إلى حد ما أكبر من الدليل وهو القيمة 1. وعلى أي حال، فإن القيمة قريبة من الواحد إلى حد يمكن أن يجعلها غير مؤثرة تأثيرا يستدعي المبادرة لعلاج.

التأثير على معاملات الانحدار

DFBETAS إحدى مقاييس تأثير المشاهدة i على كل معامل انحدار $b_k(k = 0, 1,..., p\text{-}1)$ هو الفرق بين معامل الانحدار b_k المقدَّر بالاستناد إلى المشاهدات الـ n جميعها وبين معامل الانحدار الذي نحصل عليه عند حذف المشاهدة i، وسنرمز له بـ $b_{k(i)}$ وعند قسمة هذا الفرق على معايرة مناسبة نحصل على المقياس ***DFBETAS*** .

$$(DFBETAS)_{k(i)} = \frac{b_k - b_{k(i)}}{\sqrt{MSE_{(i)} C_{kk}}} \quad k = 0,1,..., p\text{-}1 \qquad (11.27)$$

حيث c_{kk} هو العنصر القطري الـ k من $(\mathbf{X'X})^{-1}$ والمقام تقدير للخطأ المعياري لـِ b_k، حيث يُبنى MSE على n-1 من المشاهدات. والقيمة المطلقة الكبيرة لِ $(DFBETAS)_{k(i)}$ مؤشر تأثير كبير للمشاهدة i على معامل الانحدار الـ k. وكدليل لتحديد المشاهدات المؤثرة، نوصي باعتبار مشاهدة ما كمشاهدة مؤثرة إذا تجاوزت القيمة المطلقة لـِ ***DFBETAS*** الواحد في مجموعات صغيرة إلى متوسطة الحجم من البيانات، وتجاوزت $2/\sqrt{n}$ في مجموعات كبيرة من البيانات.

مثال. في مثال شحوم الجسم بمتغيرين مستقلين، نجد في الأعمدة ٢، ٣ و٤ من الجدول (١١-٣) قوائم بقيم الـ ***DFBETAS***. ونلاحظ أن المشاهدة الثالثة هي المشاهدة الوحيدة التي تتجاوز الدليل وهو القيمة لمجموعات صغيرة إلى متوسطة الحجم من البيانات، وذلك من أجل كل من b_1 و b_2. وهكذا تُوسم المشاهدة الثالثة من جديد بأنها ذات نفوذ. وعلى أي حال نقول ثانية إن قيم الـ ***DFBETAS*** لاتتجاوز الواحد تجاوزا كبيرا جدا مما يمكننا معه القول إن المشاهدة الثالثة قد لايكون لها من التأثير مايستدعي المبادرة لعلاج.

مسافة كوك. مقياس مسافة كوك D_i هومقياس إجمالي للتأثير المشترك للمشاهدة i على جميع معاملات الانحدار المقدَّرة. وهذا المقياس مستنبط من مفهوم منطقة الثقة المتزامنة لمعاملات الانحدار β_k جميعها وعددها p $(k = 0, 1 ,..., p - 1)$ ويمكن تبيان أن حدود منطقة الثقة المشتركة هذه والخاصة بنموذج الانحدار المتعدد طبيعي الأخطاء (7.18) معطاة بالعلاقة:

$$\frac{(\mathbf{b}-\beta)'\mathbf{X}'\mathbf{X}(\mathbf{b}-\beta)}{pMSE} = F(1\text{-}\alpha; p, n-p) \qquad (11.28)$$

ويستخدم مقياس مسافة كوك D_i الهيكل نفسه المستخدم لقياس التأثير المشترك للفر في معاملات الانحدار المقدّرة عند حذف المشاهدة i:

$$D_i = \frac{(\mathbf{b}-\mathbf{b}_{(i)})'\mathbf{X}'\mathbf{X}(\mathbf{b}-\mathbf{b}_{(i)})}{pMSE} \qquad (11.29)$$

حيث $\mathbf{b}_{(i)}$ متجه معاملات الانحدار المقدّرة الذي نحصل عليه عند حذف المشاهدة i كالمعتاد، هو ذاك المتجه عند استخدام المشاهدات n جميعها.

وبينما لايتبع D_i التوزيع F فقد وُجد أنه من المفيد نسبة القيمة D_i إلى التوزيع المقابل وفقا للعلاقة (11.28) ومعرفة المئين الموافق لتلك القيمة. وإذا كانت قيمة المئـ أقل من حوالي 10 أو 20 بالمائة، فيكون للمشاهدة i، على مايبدو، تأثير بسيط عل معاملات الانحدار. وعلى الوجه الآخر، إذا كانت قيمة المئين قرب الـ 50 بالمائة أكثر فينبغي إعتبار المسافة بين المتجهين $\mathbf{b}$ و $\mathbf{b}_{(i)}$ كبيرة، مما يتضمن أن للمشاهدة تأثـ كبيرا على توفيق دالة الانحدار.

ومن حسن الحظ، يمكن حساب مقياس مسافة كوك D_i دون توفيق دالة انحـ جديدة في كل مرة نحذف فيها مشاهدة مختلفة. والعبارة المكافئة جبريا هي:

$$D_i = \frac{e_i^2}{pMSE}\left[\frac{h_{ii}}{(1-h_{ii})^2}\right] \qquad (11.29a)$$

ونلاحظ من (11.29a) أن D_i يعتمد على عاملين: (١) حجم الراسب e_i و (٢) قيمـ العزم h_{ii}. وكلما كان أي من e_i أو h_{ii} أكبر. كلما كان D_i أكبر وهكذا يمكن أ تكون المشاهدة i مؤثرة : (١) إذا كان لدينا راسب e_i كبير وقيمة عزم معتدلة h_{ii} أ (٢) إذا كان لدينا قيمة عزم h_{ii} كبيرة مع راسب e_i من حجم معتدل، أو (٣) إ كان لدينا راسب e_i كبير وقيمة عزم h_{ii} كبيرة.

مثال. في مثال شحوم الجسم بمتغيرين مستقلين، يقدم العمود الخامس من الجدوا (١١-٣)، القيم D_i . ولتوضيح الحسابات سنعتبر المشاهدة 1. ونعلم من الجدوا (١١-٢) أن $e_1 = -1.683$ و $h_{11} = 0.201$. وفضلا عن ذلك،لدينا من الجدول

(٨ـ٢)جـ، $MSE = 6.47$ و $p = 3$ لنموذج بمتغيرين مستقلين. وبالتالي نجد:

$$D_1 = \frac{(-1.683)^2}{3(6.47)}\left[\frac{.201}{(1-.201)^2}\right] = .046$$

ونلاحظ من العمود الخامس في الجدول (١١-٣) أن المشاهدة الثالثة هي بوضوح المشاهدة الأكثر تأثيرا، إذ يقابلها $D_3 = 0.490$، بينما مقياس المسافة التالي في الكبر هو $D_{13} = 0.212$ وهو أصغر بكثير.

ولتثمين شأن $D_3=0.490$ ، نعود إلى التوزيع F المقابل ، ونعني $F(p,n-p)=F(3, 17)$ فيمكن تبيان أن 0.490 هي في جوار المئين 31 لهذا التوزيع. وبالتالي، يبدو أن المشاهدة الثالثة تؤثر بالفعل في توفيق الانحدار، ولكن قد لايكون مدى التأثير كبيرا إلى الحد الذي يستدعي اتخاذ تدابير علاجية.

تعليقات

١ـ يمكن النظر إلى مقياس مسافة كوك D_i على أنه يعكس، بصورة إجمالية ولكل مشاهدة، الفروق بين القيمة التوفيقية عند استخدام المشاهدات n جميعها وبين القيمة التوفيقية عند إلغاء المشاهدة i، إذ يمكن تبيان أن العبارة التالية هي عبارة مكافئة من أجل D_i:

$$D_i = \frac{(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})'(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})}{pMSE} \tag{11.30}$$

وكالمعتاد $\hat{\mathbf{Y}}$ هنا هو متجه القيم التوفيقية عندما تُستخدم المشاهدات الـ n جميعا و $\hat{Y}_i$ هو متجه القيم التوفيقية عندما تُحذف المشاهدة .

٢ـ تحليل المشاهدات القاصية والمؤثرة هو عنصر ضروري لتحليل انحدار جيد، إلا أنه ليس عملا آليا، ولا عملا مضمونا، بل يتطلب اجتهادا ناجحا من المحلل. وفي الغالب تعمل الطرق التي وصفناها بشكل طيب إلا أنها ستكون في أحيان أخرى غير فعالة. وعلى سبيل المثال، إذا كانت مشاهدتان قاصيتان مؤثرتان متطابقتين تقريبا، كما هي الحال في المشاهدتين 3 و4 المرسومتين في الشكل (١١-٤)، فإن محللا يحذف إحداهما في كل مرة ويقدر التغير في التوفيق سيجد في النتيجة أن لاتغيّر لكل من

هاتين المشاهدتين القاصيتين. وسبب ذلك هو أن المشاهدة القاصية التي احتفظ بها ستحجب تأثير المشاهدة القاصية المحذوفة.

التأثير على الاستقراءات

واستكمالا لتحديد المشاهدات المؤثرة نقول إنها فكرة جيدة في العادة أن نقوم بطريقة مباشرة بفحص الاستقراءات من نموذج الانحدار التوفيقي التي يمكن استنباطها مع وبدون المشاهدة (أو المشاهدات) المعنية. وإذا لم تتغير الاستقراءات تغيرا جوهريا فهناك القليل من الحاجة للتفكير في تدابير علاجية تتعلق بالمشاهدات التي شُخِّصَت بأنها مؤثرة. وعلى الوجه الآخر، فإن ترافق إلغاء المشاهدة بتغيرات جدية في الاستقراءات المستنبطة من النموذج التوفيقي ستستدعي التفكير في تدابير علاجية وسنناقش في الفقرة القادمة بعض التدابير العلاجية الممكنة.

مثال. في مثال شحوم الجسم بمتغيرين مستقلين، حددنا المشاهدتين 3 و 5 كمشاهدتين قاصيتين في قيم X والمشاهدتين 8 و 13 كمشاهدتين قاصيتين في قيم Y وجميع مقاييس التأثير الثلاثة (***DFFITS*** ، ***DFBETAS***، ومسافة كوك) حددت المشاهدة الثالثة فقط كمشاهدة مؤثرة، واقترحت، في الحقيقة، أن أهمية تأثيرها قد تكون هامشية بحيث لاتستدعي تدابير علاجية.

ولكن المحلل في مثال شحوم الجسم كان يهتم في المقام الأول بتوفيق نموذج الانحدار، لأن الغرض من النموذج كان استخدامه للقيام بتنبؤات ضمن مدى المشاهدات على المتغيرات المستقلة في مجموعة البيانات، وبالتالي فقد اعتبر المحلل دالتي الانحدار التوفيقيتين مع المشاهدة 3 وبدونها:

مع المشاهدة 3 $\qquad \hat{Y} = -19.174 + .2224X_1 + .6594X_2$

بدون المشاهدة 3 $\qquad \hat{Y} = -12.428 + .5641X_1 + .3635X_2$

وبسبب الخطية المتعددة المرتفعة بين X_1 و X_2، لم يُفاجأ المحلل بالتغيرات في مقداري b_1 و b_2 عندما أُلغيت المشاهدة. فلنتذكر أن الانحرافين المعياريين المقدَّرين للمعاملين، وهما معطيان في الجدول (٨-٢)جـ، كبيران جدا، وأن مشاهدة واحدة يمكن أن تغيّر المعاملين المقدَّرين تغييرا كبيرا جدا عندما يكون المتغيران مرتبطين ارتباطا عاليا.

ولفحص تأثير المشاهدة الثالثة على استقراءات سيقوم بها المحلل بطريقة مباشرة في مدى المشاهدات X مستخدما دالة الانحدار التوفيقية، فقد قام بحساب الفرق بين القيمة التوفيقية $\hat{Y}_i$ المستندة إلى المشاهدات العشرين جميعها والقيمة التوفيقية $\hat{Y}_{i(3)}$ الناتجة عند الغاء المشاهدة الثالثة، وذلك من أجل كل من المشاهدات العشرين. والمقياس المعنيّ كان متوسط مطلق الفروق النسبية المئوية:

$$\frac{\sum_{i=1}^{n}\left|\frac{\hat{Y}_{i(3)}-\hat{Y}_i}{\hat{Y}_i}\right|100}{n}$$

ومتوسط مطلق الفروق هذا كان 3.1 بالمائة، وفضلا عن ذلك فإن 17 من الفروق العشرين كانت أقل من 5 بالمائة (الحسابات غير مبينة). وعل أساس هذه البينة المباشرة حول تأثير المشاهدة على الاستقراءات التي ستتّخذ، اقتنع المحلل بأن المشاهدة الثالثة لاتمارس نفوذا مفرطا مما لايستدعي القيام بمعالجة هذه الحالة.

(١١–٥) تدابير علاجية لمشاهدات مؤثرة

بعد استخدام مقاييس مصفوفة القبّعة، رواسب الحذف المعيّرة تقديرا، ***DFFITS***، ***DFBETAS***، ومسافة كوك، لتحديد المشاهدات القاصية المؤثرة التي يكون لها أثر كبير على انحدار المربعات الدنيا التوفيقي، يجب تقرير مايمكن عمله بالنسبة لمشاهدات كهذه. ومن الواضح أنه لاينبغي نبذ مشاهدة قاصية مؤثرة بصورة آلية، ذلك لأنها يمكن أن تكون صحيحة تماما وتمثل ببساطة حادثة غير محتملة. وقد يؤدي نبذ مشاهدة قاصية إلى نتيجة غير مرغوبة هي زيادة تباينات بعض معاملات الانحدار المقدّرة.

وعلى الوجه الآخر، إذا كانت الظروف المحيطة بالبيانات تقدم تفسيرا للمشاهدة غير العادية يشير إلى حالة استثنائية لايهدف النموذج إلى تغطيتها فقد يكون من المناسب نبذ المشاهدة.

وعندما تكون المشاهدة القاصية المؤثرة دقيقة فقد لاتمثل حادثة غير محتملة بل تمثل فشلا للنموذج. وقد يكون الفشل في إلغاء متغير مستقل مهم، أو اختيار شكل دالي غير صحيح مثل إلغاء تأثير منحن لمتغير مستقل شمله النموذج، أو إلغاء حد تفاعل

مهم. وفي الغالب، يقود تحديد المشاهدات القاصية المؤثرة إلى بصيرة نافذة لها آثارها القيّمة في مجال تقوية النموذج.

وعندما تكون مشاهدة قاصية مؤثرة دقيقة ولكن لايمكن إيجاد تفسير لها فإن البديل الأقل قسوة من نبذها هو تلطيف آثارها. وإحدى الوسائل الممكنة لتلطيف الآثار هو استخدام تحويل. وعلى سبيل المثال، إذا كانت مشاهدة قاصية في أحد المتغيرات X فقد يجلب تحويل مثل التحويل اللوغاريتمي أو تحويل الجذر التربيعي المشاهدة القاصية لتصبح أقرب إلى بقية المشاهدات وبذلك تخف آثارها. وبالطبع، يحتاج المرء إلى التحقق من استمرار صلاحية النموذج مع وجود المتغير الجديد بعد التحويل وذلك للتأكد من أن التحويل لم يخلق بدوره مشاكل جديدة. والوسائل الممكنة الأخرى لتلطيف الأثر هو استخدام طريقة تقدير مختلفة. وسنناقش الآن واحدة من طرق التقدير البديلة هذه.

طريقة الانحرافات المطلقة الدنيا

هذه الطريقة هي واحدة من طرق منيعة متنوعة تمتلك خاصة أنها غير حساسة لكل من قيم قاصية في البيان الإحصائي أو لعيوب في صلاحية النموذج المستخدم. وتقدّر طريقة الانحرافات المطلقة الدنيا معاملات الانحدار بجعل مجموع القيم المطلقة لانحرافات المشاهدات عن متوسطاتها أصغر مايمكن. والمعيار الذي نريد جعله أصغر مايمكن هو:

$$\sum_{i=1}^{n}\left|Y_i-(\beta_0+\beta_1X_{i1}+\ldots+\beta_{p-1}X_{i,p-1})\right| \tag{11.31}$$

وبما أن القاعدة هنا تنطوي على الانحرافات المطلقة بدلا من مربعاتها، فإن طريقة الانحرافات المطلقة الدنيا تضع من التأكيد على المشاهدات القاصية أقل مما تضعه طريقة المربعات الدنيا.

ويمكن الحصول على معاملات الانحدار المقدّره وفقا لطريقة الانحرافات المطلقة الدنيا باستخدام تقانات البرمجة الخطية. ويمكن العثور على تفاصيل تتعلق بالنواحي الحسابية في كتب مدرسية مختصة، مثل المرجع [11.3].

مثال. في مثال شحوم الجسم بمتغيرين مستقلين، وُسمت المشاهدة الثالثة بأنها ذات تأثير كبير على دالة الانحدار التوفيقية ورأينا أن دالتي الانحدار التوفيقيتين مع المشاهدة الثالثة وبدونها كانتا:

مع المشاهدة 3 $\hat{Y} = -19.174 + .2224X_1 + .6594X_2$

بدون المشاهدة 3 $\hat{Y} = -12.428 + .5641X_1 + .3635X_2$

ومع استخدام طريقة الانحرافات المطلقة الدنيا، نجد دالة الانحدار التوفيقية:

$$\hat{Y} = -17.027 + .4173X_1 + .5203X_2$$

وهكذا نرى أن طريقة الانحرافات المطلقة الدنيا تؤدي إلى تغيرات أكثر تواضعاً مما يؤدي إليه إسقاط المشاهدة 3 بالكامل. ويبين تحليل الرواسب أن طريقة الانحرافات المطلقة الدنيا تنتج تخفيضات في تلك الرواسب التي كانت مع طريقة المربعات الدنيا أكبر في قيمتها المطلقة.

تعليقات

١- سوف لا تجمع الرواسب لطريقة الانحرافات المطلقة الدنيا، في العادة إلى الصفر.

٢- قد لا يكون الحل لمعاملات الانحدار المقدَّرة فريدا في طريقة الانحرافات الدنيا المطلقة.

٣- تدعى طريقة الانحرافات المطلقة الدنيا أيضا الانحرافات المطلقة الصغرى، والمجموع الأصغري للانحرافات المطلقة، والعمودي L_1 الأصغري.

٤- اقترح العديد من الطرق المنيعة الأخرى إلى جانب طريقة الانحرافات المطلقة الدنيا، ويناقش المرجع [11.4] عددا من هذه الطرق.

(١١-٦) تشخيصات الخطية المتعددة - عامل تضخم التباين

عندما ناقشنا الخطية المتعددة في الفصل الثامن، ذكرنا بعض المشاكل الرئيسة التي تبرز بصورة تقليدية عندما تكون المتغيرات المستقلة المعتبرة في نموذج الانحدار مرتبطة فيما بينها ارتباطا عاليا:

١- إضافة أو حذف متغير مستقل يغير معاملات الانحدار.

٢- يتغير مجموع المربعات الإضافي المترافق مع متغير مستقل، معتمدا على أيّ المتغيرات المستقلة الأخرى مشمول في النموذج.

٣- تصبح الانحرافات المعيارية المقدرة لمعاملات الانحدار كبيرة عندما تكون المتغيرات المستقلة في نموذج الانحدار مرتبطة فيما بينها ارتباطا عاليا.

٤- قد لاتكون معاملات الانحدار المقدَّرة كل بمفردها مهمة إحصائيا مع أن هناك بصورة حاسمة علاقة إحصائية بين المتغير التابع والمتغيرات المستقلة.

ويمكن أن تنشأ هذه المشاكل أيضا دون وجود درجة عالية من الخطية المتعددة، ولكن فقط تحت ظروف غير عادية وغير محتملة في التطبيق العملي.

وسنعرض الآن بعض التشخيصات غير الرسمية للخطية المتعددة بالإضافة إلى تشخيص رسمي مفيد للغاية هو عامل تضخّم التباين.

تشخيصات غير رسمية

تقدم التشخيصات غير الرسمية التالية مؤشرات لوجود خطية متعددة جدية:

١- تغيرات كبيرة في معاملات الانحدار المقدَّرة عند إضافة أو حذف متغير، أو عند تعديل أو حذف مشاهدة.

٢- نتائج غير معنوية في اختبارات فردية حول معاملات الانحدار الخاصة بمتغيرات مستقلة مهمة.

٣- معاملات انحدار مقدَّرة، إشارتها الجبرية معاكسة تماما لما تتوقعه الاعتبارات النظرية، أو الخبرة السابقة.

٤- معاملات كبيرة للارتباط البسيط بين أزواج من المتغيرات المستقلة في مصفوفة الارتباط r_{XX}.

٥- فترات ثقة عريضة لمعاملات انحدار تمثل متغيرات مستقلة مهمة.

مثال. نعتبر من جديد مثال شحوم الجسم في الفصل الثامن، وهذه المرة بالمتغيرات المستقلة الثلاثة جميعها – سماكة جلد العضلة ثلاثية الرؤوس X_1، محيط الفخذ X_2،

ومحيط منتصف الذراع X_3. وقد أعطي البيان الإحصائي في الجدول (٨-١) وذكرنا هناك أن المتغيرين المستقلين: سماكة جلد العضلة ثلاثية الرؤوس ومحيط الفخذ مرتبطان ارتباطا عاليا. ولاحظنا أيضا في الفصل الثامن تغيرات في معاملات الانحدار المقدَّرة وفي أخطائها المعيارية المقدَّرة عند إضافة متغير، ونتائج غير معنوية في اختبارات بمفردها حول متغيرات يُنتظَر أن تكون متغيرات مهمة، ومعامل مقدَّر سالب حيث يُتوقَّع أن يكون موجبا. وهذه جميعها مؤشرات غير رسمية تقترح وجود خطية متعددة جدية بين المتغيرات المستقلة.

ملاحظة

للطرق غير الرسمية التي وصفناها لتوّنا آفاق محدودة، فهي لاتقدّم مقاييس كمية لزخم الخطية المتعددة، ولايمكن لها أن تحدد طبيعتها. وعلى سبيل المثال، إذا كان الارتباط بين أزواج المتغيرات X_1، X_2 و X_3 منخفضا فإن فحص معاملات الارتباط البسيطة سوف لا يفصح بالضرورة عن وجود علاقات بين مجموعات من المتغيرات المستقلة.

والمحدودية الأخرى لطرق التشخيص غير الرسمية هي إمكانية وقوع المسلك الملحوظ أحيانا دون وجود خطية متعددة.

عامل تضخم التباين

إحدى الطرق الرسمية المستخدمة على نطاق واسع للكشف عن وجود خطية متعددة هي استخدام عوامل تضخّم التباين، وتقيس هذه العوامل مدى تضخم تباينات معاملات الانحدار المقدَّرة بالمقارنة مع حالة عدم وجود صلة خطية بين المتغيرات المستقلة.

ولفهم أهمية عوامل تضخم التباين، نبدأ بدقة معاملات الانحدار المقدّرة بطريقة المربعات الدنيا، وهذه الدقة تقاس بالتباينات. ونعلم من (7.41) أن مصفوفة التباين ـ التغاير لمعاملات الانحدار المقدّرة هي:

$$\sigma^2\{\mathbf{b}\} = \sigma^2(\mathbf{X'X})^{-1} \tag{11.32}$$

ولتخفيف أخطاء التدوير في حسابات $(\mathbf{X'X})^{-1}$، ذكرنا في الفصل الثامن أنه من المستحسن البدء بتحويل المتغيرات باستخدام تحويل الارتباط (8.41). وعند توفيق

النموذج بعد التحويل (8.42) فإن معاملات الانحدار المقدّرة b'_k هي معاملات معيّرة ترتبط بمعاملات الانحدار المقدَّرة للمتغيرات الأصلية قبل التحويل وفقـا للعلاقـة (8.50) ومن (11.32) يمكن الحصول على مصفوفة التباين ـ التغاير لمعاملات الانحـدار المعياريـة المقدّرة، بعد استخدام النتيجة في (8.47) التي تقول إن المصفوفـة $\mathbf{X'X}$ للمتغـيرات بعـد التحويل هي $\mathbf{r}_{XX}$.

$$\sigma^2\{b\} = (\sigma')^2\, r_{XX}^{-1} \qquad (11.33)$$

حيث $\mathbf{r}_{XX}$ مصفوفة معاملات الارتباط البسيط بين أزواج المتغيرات X، كما عرفناها في (8.44)، و $(\sigma')^2$ هو تباين حد الخطأ في النموذج بعد التحويل.

ونلاحظ من (11.33) أن تباين $b'_k\ (k = 1, \dots, p-1)$ يساوي جداء تباين حد الخطأ $(\sigma')^2$ بالعنصر القطري الـ k من المصفوفة $\mathbf{r}_{XX}^{-1}$. ويدعى هذا العامل الثاني عامل تضخم التباين (VIF)، ويمكن تبيان أن عامل تضخم التباين لـ b'_k، ونرمز له بـ $(VIF)_k$ هو:

$$(VIF)_k = (1 - R_k^2)^{-1} \qquad k = 1,2,\dots,p-1 \qquad (11.34)$$

حيث R_k^2 معامل التحديـد المتعـدد عنـد انحـدار X_k علـى الــ $p-2$ مـن المتغـيرات X الأخرى في النموذج. وبالتالي لدينا:

$$\sigma^2\{b'_k\} = (\sigma')^2 (VIF)_k = \frac{(\sigma')^2}{1 - R_k^2} \qquad (11.35)$$

وقد قدمنا في (8.62) النتائج الخاصة بـ $\sigma^2\{b'_k\}$ عندما يكون $p - 1 = 2$، وفي هـذه الحالة يكون R_k^2 مساويا لـ r_{12}^2 معامل التحديد البسيط بين X_1 و X_2.

وعامل تضخم التباين $(VIF)_k$ يساوي الواحد عندما $R_k^2 = 0$ ، أي عندما لايكون X_k علـى صلـة خطيـة بالمتغـيرات المستقلة الأخـرى. وإذا كـان $R_k^2 \neq 0$ فـإن $(VIF)_k$ عندئذ أكبر من الواحد مشيرا إلى تضخم في تباين b'_k . وهـذا واضـح مـن (11.35) إذ يصبح المقام أصغر كلما أصبحت R_k^2 أكبر، مما يؤدي إلى تباين أكـبر وعندمـا تكـون X_k على علاقة خطية تامة بالمتغيرات المسـتقلة الأخـرى في النمـوذج ممـا يجعـل $R_k^2 = 1$ فعندئذ يكون $(VIF)_k$ و $\sigma^2\{b'_k\}$ غير محدودين.

استخدامات تشخيصية. غالبا ماتُستخدم أكبرقيمة *VIF* من بين المتغيرات X جميعها كمؤشر لمدى خطورة الخطية المتعددة وفي الغالب تُتخذ قيمة عظمى لـ *VIF* تتجاوز العشرة كمؤشر إلى إمكانية تأثير غير مقبول للخطية المتعددة على تقديرات المربعات الدنيا.

ويقدم متوسط القيم *VIF* معلومات عن خطورة الخطية المتعددة بدلالة المسافة بين معاملات الانحدار المعيارية المقدّرة b'_k وبين القيم الحقيقية β'_k. ويمكن تبيان أن القيمة المتوقعة لمجموع مربعات الخطأ هذه $(b'_k - \beta'_k)^2$ معطاة بالعلاقة:

$$E\left\{\sum_{k=1}^{p-1}(b'_k - \beta'_k)^2\right\} = (\sigma')^2 \sum_{k=1}^{p-1}(VIF)_k \qquad (11.36)$$

وهكذا فإن قيما كبيرة لِ *VIF* تنتج، في المتوسط، فروقا أكبر بين معاملات الانحدار المعيارية المقدَّرة ومعاملات الانحدار المعيارية الحقيقية.

وعندما لا توجد صلة خطية لأي متغير X بالمتغيرات المستقلة الأخرى في نموذج الانحدار، أي $R_k^2 = 0$ فعندئذ، $(VIF)_k \equiv 1$ و:

$$E\left\{\sum_{k=1}^{p-1}(b'_k - \beta'_k)^2\right\} = (\sigma')^2 (p-1) \qquad ,(VIF)k \equiv 1 \qquad (11.36a)$$

وتقدم نسبة النتيجتين في (11.36) و(11.36a) معلومات مفيدة عن تأثير الخطية المتعددة على مجموع مربعات الخطأ:

$$\frac{(\sigma')^2 \sum(VIF)_k}{(\sigma')^2 (p-1)} = \frac{\sum(VIF)_k}{p-1}$$

ونلاحظ أن هذه النسبة هي ببساطة، متوسط قيم الـ *VIF* وسنرمز لها بـ $\overline{(VIF)}$:

$$\overline{(VIF)} = \frac{\sum_{i=1}^{p-1}(VIF)_k}{p-1} \qquad (11.37)$$

ومتوسط قيم *VIF* أكبر بكثير من الواحد هو مؤشر لمشاكل جدية للخطية المتعددة.

مثال. يتضمن الجدول (١١-٤) معاملات الانحدار المعيارية المقدّرة وقيم *VIF* لمثال شحوم الجسم بثلاث متغيرات مستقلة (الحسابات غير مبينة). وقيمة *VIF* العظمى هي 708.84 و $\overline{(VIF)} = 459.26$. وهكذا فإن مجموع مربعات الخطأ المتوقع في معاملات انحدار المربعات الدنيا المعيارية أكبر بما يقرب من 460 مرة منها لوكانت المتغيرات X

غير مرتبطة. وبالإضافة إلى ذلك، فإن جميع قيم *VIF* الثلاث تتجاوز العشرة بكثير، مما يشير إلى وجود مشاكل جدية للخطية المتعددة.

ومن المفيد ملاحظة أن $(VIF)_3 = 105$ بالرغم من حقيقة أن كـلا مـن r_{13}^2 و r_{23}^2 (انظر الجدول (٨-١)ب) غير كبيرين. وهاهنا مثال يكون فيه X_3 على صلة قوية بـ X_1 و X_2 معا $(R_3^2 = 0.990)$ مع أن معـاملات التحديـد البسـيط مثنـى مثنـى ليسـت كبيرة. وفحص الارتباطات الثنائية لايكشف عن هذه الخطية المتعددة.

الجدول (١١-٤) عوامل تضخم التباين لمثال شحوم الجسم بثلاثة متغيرات مستقلة

متغير	b_k'	$(VIF)k$
X_1	4.2637	708.84
X_2	-2.9287	564.34
X_3	-1.5614	104.61
$(VIF)k = 708.84$		$\overline{(VIF)} = 459.26$

تعليقات

١- يستخدم عدد من برامج الانحـدار الحاسـوبية مقلـوب عـامل تضخـم التبـاين لكشف حالات ينبغي ألا نسمح فيهـا لمتغـير X بـالدخول إلى نمـوذج انحـدار توفيقـي بسبب فرط ارتفاع تبعية ضمنية لهذا المتغير بالمتغيرات X الأخرى في النموذج. وحدود التساهل لـ $1/(\mathbf{VIF})_k = 1 - \mathbf{R}_k^2$ المستخدمة بكثرة هـي 0.01 و 0.001 أو 0.0001 فـدون هذه الحدود لايدخل المتغير إلى النموذج.

٢- محدودية عوامل تضخم التباين في كشف الخطيات المتعددة هي أنها لاتستطيع التمييز بين عدة خطيات متعددة متواقتة في آن واحد.

٣- اقتُرح عدد من الطرق الرسمية الأخرى للكشف عن خطية متعـددة. وهـي أكثر تعقيـدا من عوامل تضخم التباين، وقد نوقشت في كتب مدرسية متخصصة مثل المرجع [11.5].

(١١-٧) تدابير علاجية للخطية المتعددة ـ انحدار الحافة

نعتبر الآن بعـض التدابـير العلاجيـة لخطيـة متعـددة خطـيرة، وهـي تدابـير يمكـن استخدامها مع طريقة المربعات الدنيا المعتادة، ومن ثم نبدأ بمناقشة انحدار الحافة، وهـي طريقة في التغلب على مشاكل خطية متعددة خطرة تلجـأ إلى تعديـل طريقـة المربعـات الدنيا.

تدابير علاجية تُستخدم مع طريقة المربعات الدنيا

١- وكما رأينا في الفصل الثامن، ففي الغالب لايؤثر وجود خطية متعددة خطرة في فائدة النموذج التوفيقي في القيام باستقراءات حول متوسط الاستجابة أو القيام بتنبؤات، شريطة أن تتبع قيم المتغيرات المستقلة التي ستتناولها الاستقراءات نمط الخطية المتعددة نفسه الذي تتبعه البيانات التي بُني عليها نموذج الانحدار. وبالتالي يكون أحد التدابير العلاجية هو أن يقتصر استخدام نموذج الانحدار التوفيقي على استقراءات حول قيم للمتغيرات المستقلة تتبع نمط الخطية المتعددة نفسه.

٢- وكما لاحظنا في نماذج انحدار كثيرات الحدود، في الفصل التاسع، فإن التعبير عن المتغير، أو المتغيرات المستقلة، في صيغة انحرافات عن المتوسط يخدم في تخفيض الخطية المتعددة بين حدود من المرتبة الأولى والمرتبة الثانية والمراتب الأعلى، لأي متغير مستقل، تخفيضا كبيرا.

٣- يمكن إسقاط متغير واحد، من بين عدة متغيرات مستقلة، للتقليل من الخطية المتعددة، وبالتالي تخفيض الأخطاء المعيارية لمعاملات الانحدار المقدَّرة للمتغيرات المستقلة التي بقيت في النموذج. ولهذا التدبير العلاجي محدوديتين مهمتين. فأولا، لانحصل على أية معلومات مباشرة عن المتغيرات المستقلة التي أسقطناها. وثانيا، تتأثر مقادير معاملات الانحدار للمتغيرات المستقلة الباقية في النموذج بالمتغيرات المستقلة المرتبطة معها والتي لم يشملها النموذج.

٤- يمكن أحيانا إضافة بعض المشاهدات التي تكسر نمط الخطية المتعددة، إلا أن هذا الاختيار لايتوافر، في الغالب. ففي التجارة والاقتصاد، مثلا، هناك العديد من المتغيرات المستقلة التي لايمكن التحكم فيها، وبالتالي ستميل المشاهدات الجديدة إلى إظهار أنماط الخطية المتعددة نفسها التي أظهرتها المشاهدات السابقة.

٥- في بعض الدراسات الاقتصادية، يمكن تقدير معاملات الانحدار، لمتغيرات مستقلة مختلفة، من مجموعات بيانات مختلفة. وذلك لتجنب مشاكل الخطية المتعددة. ولهذه الغاية، يمكن أن تستخدم دراسات الطلب، مثلا، بيانات متقاطعة عرضيا أو

بيانات سلسلة زمنية. فلنفرض أن المتغيرات المستقلة في دراسة طلب هي السعر والدخل، والعلاقة المراد تقديرها هي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad (11.38)$$

حيث Y الطلب، X_1 الدخل و X_2 السعر. عندئذ يمكن تقدير معامل الدخل β_1 من بيانات متقاطعة عَرْضيا. وبالتالي نعدّل متغير الطلب Y:

$$\mathbf{Y}_i' = \mathbf{Y}_i - \mathbf{b}_1 \mathbf{X}_{i1} \qquad (11.39)$$

وأخيرا نقدّر معامل السعر β_2 من انحدار متغير الطلب المعدَّل Y' على X_2.

انحدار الحافة

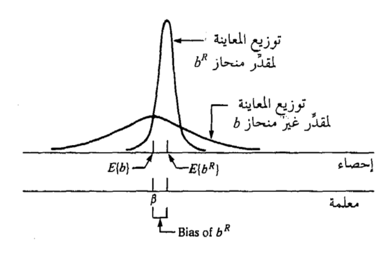
التقدير المنحاز. انحدار الحافة هو واحد من عدة طرق اقتُرِحت لعلاج مشاكل الخطية المتعددة وذلك بتعديل طريقة المربعات الدنيا بحيث تسمح بمقدّرات منحازة لمعاملات الانحدار. وعندما ينحاز مقدِّر بمقدار بسيط فقط ويكون أكثر دقة بكثير من مقدِّر غير منحاز فقد يكون المقدر المفضل، لأن احتمال قربه من القيمة الحقيقية للمعلمة سيكون، عندئذ، احتمالا أكبر. ويوضح الشكل (١١-٧) هذه الحالة. فالمقدِّر b غير منحاز ولكنه غير دقيق، بينما المقدر b^R أكثر دقة بكثير ولكنه منحاز انحيازا بسيطا. واحتمال وقوع b^R قرب القيمة الحقيقية β أكبر بكثير مما هو في حالة المقدِّر غير المنحاز b.

والقيمة المتوقعة لمربع انحراف المقدّر المنحاز b^R عن المعلمة الحقيقية β هي قياس للتأثير المركب للانحياز وتغيّر المعاينة. ويدعى هذا القياس متوسط مربعات الخطأ، ويمكن تبيان أنه يساوي:

$$E\{b^R - \beta\}^2 = \sigma^2\{b^R\} + (\{E\{b^R\} - \beta)^2 \qquad (11.40)$$

وهكذا يكون متوسط مربعات الخطأ مساويا لتباين المقدّر مضافا إليه مربع الانحياز. ونلاحظ تطابق متوسط مربعات الخطأ وتباين المقدّر إذا كان المقدّر غير منحاز.

شكل (١١-٧) قد يكون مقدّر منحاز مع تباين صغير مفضلا على مقدّر غير منحاز مع تباين كبير



مقدرات الحافة. تعطى المعادلات الناظمية للمربعات الدنيا المعتادة بالعلاقة

$$(\mathbf{X'X})\mathbf{b} = \mathbf{X'Y} \tag{11.41}$$

وعند تحويل المتغيرات جميعها وفقا لتحويل الارتباط (8.41) يُعطى نموذج الانحدار بعد التحويل بالعلاقة (8.42):

$$Y_i' = \beta_1' X_{i1}' + \beta_2' X_{i2}' + \ldots + \beta_{p-1}' X_{i,p-1}' + \varepsilon_i' \tag{11.42}$$

وتعطى المعادلات الناظمية للمربعات الدنيا بالعلاقة (8.49a):

$$\mathbf{r}_{XX}\mathbf{b} = \mathbf{r}_{YX} \tag{11.43}$$

حيث $\mathbf{r}_{XX}$ مصفوفة ارتباط المتغيرات X المعرفة (8.44) و $\mathbf{r}_{XX}$ هو متجه معاملات الارتباط البسيط بين Y وكل متغير من المتغيرات X، وهذا المتجه معرف في (8.45).

ونحصل على مقدّرات انحدار الحافة المعيّر بإدخال ثابت انحياز غير سالب $c \geq 0$ إلى المعادلات الناظمية للمربعات الدنيا (11.43)، وذلك بالصيغة التالية:

$$(\mathbf{r}_{XX} + c\mathbf{I})\mathbf{b}^R = \mathbf{r}_{YX} \tag{11.44}$$

حيث $\mathbf{b}^R$ متجه معاملات انحدار الحافة المعياري b_k^R:

$$\underset{(p-1)\times 1}{\mathbf{b}^R} = \begin{bmatrix} b_1^R \\ b_2^R \\ \cdot \\ \cdot \\ \cdot \\ b_{p-1}^R \end{bmatrix} \tag{11.45}$$

و I هي $(p - 1) \times (p - 1)$ مصفوفـة وحـدة. ويُنتـج حـل المعـادلات الناظميـة (11.44) معادلات انحدار الحافة المعياري:

$$\mathbf{b}^R = (\mathbf{r}_{XX} + c\mathbf{I})^{-1}\mathbf{r}_{YX} \qquad (11.46)$$

ويعكس الثابت c مقدار الانحيـاز في المقـدِّرات. وعندمـا يكـون $c = 0$ تُختـزل (11.46) إلى معاملات انحدار المربعات الدنيا المعتادة في صيغة معيارية، كما رأيناهـا في (8.49b)، وعندما يكون $c > 0$ فإن معاملات انحدار الحافة تكون منحازة ولكنها تميـل إلى أن تكون أكثر استقرارا (أي أقل تغيرا) من مقدرات المربعات الدنيا المعتادة.

اختيار ثابت الانحياز c. يمكن تبيان أن مركبة الانحيـاز لمتوسـط مربعـات الخطـأ الاجمالي لمقدّر انحدار الحافة b^R تزداد بازدياد c (مع انتهاء جميع المقدرات b_k^R في اتجـاه الصفر)، بينما يصبح تباين المركبة، في الوقت نفسه، أصغر. وفضلا عـن ذلـك، يمكـن تبيان أنه توجد دائما قيمة ما c يكون لمقدّرات انحدار الحافـة $\mathbf{b}^R$ مـن أجلهـا، متوسـط مربعات خطأ إجمالي أصغر مما هو لمقدّرات المربعات الدنيا المعتادة $\mathbf{b}$. وتكمن الصعوبـة في أن القيمة المثلى لِـ c تتغير من تطبيق إلى آخر، وهي غير معروفة.

وتستند طريقة شائعة الاسـتخدام لتحديـد ثـابت الانحيـاز c إلى مـا يُسـمى بـأثر الحافة وإلى عوامل تضخـم التبـاين $(VIF)_k$ المعطـاة في (11.34) وأثـر الحافـة هـو رسـم متزامن لقيم المعاملات المقدّرة لانحدار الحافة المعياري، وعددها $p - 1$، وذلك من أجـل قيم مختلفة لِـ c تقع عادة بين الصفر والواحد. وتشير الخبرة الواسعة إلى إمكانية تذبذب معامل الانحدار المقدّر b_k^R تذبذبا واسعا عندما تنزاح c، قليلا عن القيمة صفر، لا بـل يمكن أن تغير إشارتها. إلا أن هذه التذبذبات الواسعة تتوقف، تدريجيـا، ويميـل مقـدار معامل الانحدار إلى التغير تغيرا بطيئا فقط عندما يزداد c شيئا فشيئا. وفي الوقت نفسـه تميل قيمة الـ $(VIF)_k$ إلى الهبوط بسرعة عندما تنزاح c قليلا عن الصفر، وتميل قيمة الـ $(VIF)_k$ بصورة تدريجية أيضا، إلى مجرد التغير باعتدال عند زيادة c شيئا فشيئا. ولذلك يقوم المحلل بفحـص أثـر الحافـة وقيـم الـ VIF، ويختـار أصغـر قيمـة لـِ c تبـدو معهـا معاملات الانحدار وكأنها بدأت تستقر وللمرة الأولى في أثر الحافة، وتصبح معها قيـم الـ VIF صغيرة صغرا كافيا. وهكذا فإن الاختيار هو مسألة اجتهاد.

مثال. لاحظنا سابقا عدة مؤشرات غير رسمية لخطية متعـددة شـديدة في بيانـات مثـال شحوم الجسم بثلاثة متغيرات مستقلة. وفي الحقيقـة، فـإن معـامل الانحـدار المقـدر b_2 سالب في النموذج التوفيقي بثلاثة متغيرات مستقلة (جدول (٨ـ٢) د) مـع أنـه مـن المتوقع أن يرتبط مقدار شحوم الجسم إيجابا مع محيط الفخـذ. وقـد أُجريـت حسـابات انحدار الحافة لبيانات مثال شحوم الجسـم في الجـدول (٨ـ١) (الحسـابات غـير معطـاة هنا). ومعاملات انحدار الحافة المعياري مقدمة في الجدول (١١ـ٥) من أجل قيم مختـارة لِـ c، وعوامل تضخم التباين معطاة في الجدول (١١ـ٦)، الذي يتضمن أيضا معاملات التحديد المتعدد R^2 ويقدم الشكل (١١ـ٨) أثر الحافة لمعاملات الانحدار المعياري المقـدّرة. وتسهيلا للتحليل. فإن السلّم الأفقي لـ c في الشكل (١١ـ٨) هو سلّم لوغاريتمي.

لاحظ عدم الاستقرار في الشكل (١١ـ٨) لمعاملات الانحدار من أجـل قيـم جـد صغيرة لـ c، إذ يغير معامل الانحـدار المقـدَّر b_2^R، في الحقيقـة، إشـارته. ولاحـظ أيضـا سرعة التناقص السريع في قيم VIF في الجدول (١١ـ٦). وقد تقرّر استخدام $c = 0.02$ هنا، لأن لمعاملات انحدار الحافة قيما VIF قريبـة مـن الواحـد، مـن أجـل هـذه القيمـة لثابت الانحياز، وتبدو معاملات الانحدار المقدّرة وقد أصبحت مستقرة بصورة معقولة. والنموذج التوفيقي الناتج من أجل $c = 0.02$ هو:

$$\hat{Y}' = 0.5463X_1' + 0.3774X_2' - 0.1369X_3'$$

وبالعودة إلى المتغيرات الأصلية وفقا لـ (8.50) نجد:

$$\hat{Y} = -7.3978 + 0.5553X_1 + 0.3681X_2 - 0.1917X_3$$

حيث:

$$\bar{Y} = 20.195 \text{ ،} \quad \bar{X}_1 = 25.305 \text{ ،} \quad \bar{X}_2 = 51.170 \text{ ،} \quad \bar{X}_3 = 27.620 \text{ ،}$$

$$s_Y = 5.106 \text{ ،} \quad s_1 = 5.023 \text{ ،} \quad s_2 = 5.235 \text{ و} \quad s_3 = 3.647$$

وقـد أُلغيـت الآن الإشـارة غـير المناسـبة لتقدير β_2، وتتسـق معـاملات الانحـدار المقدّرة اتســاقا أفضـل مـع التوقعـات المسبقة. وقـد ازداد مجمـوع مربعـات الرواسـب للمتغيرات بعد التحويل، وهو يزداد مع c، من 0.1986 عنـد $c = 0$ إلى 0.2182 عنـد $c = 0.02$، بينما تناقص R^2 من 0.8014 إلى 0.7818. وهذه التغيرات متواضعـة نسـبيا. والمتوسط المقدّر لشحوم الجسم عندما يكون $X_{h1} = 25.0$، $X_{h2} = 50.0$ و $X_{h3} = 29.0$

هو 19.33 وذلك في انحدار الحافة عند $c = 0.02$ مقارنة مع 19.19 عند استخدام حلول المربعات الدنيا المعتادة. وهكذا يبدو حل الحافة عند $c = 0.02$ مرضيا تماما هنا، ويشكل بديلا لحل المربعات الدنيا المعتادة.

جدول (١١-٥) المعاملات المقدّرة لانحدار الحافة المعياري من أجل ثوابت انحياز مختلفة مثال شحوم الجسم بثلاثة متغيرات مستقلة.

b_3^R	b_2^R	b_1^R	c
-1.561	-2.929	4.264	0.00
-.7087	-.9408	2.035	.001
-.4813	-.4113	1.441	.002
-.3758	-.1661	1.165	.003
-.3149	-.0248	1.006	.004
-.2751	.0670	.9028	.005
-.2472	.1314	.8300	.006
-.2264	.1791	.7760	.007
-.2103	.2158	.7343	.008
-.1975	.2448	.7012	.009
-.1870	.2684	.6742	.010
-.1369	.3774	.5463	.020
-.1181	.4134	.5004	.030
-.1076	.4302	.4760	.040
-.1005	.4392	.4605	.050
-.0952	.4443	.4494	.060
-.0909	.4472	.4409	.070
-.0873	.4486	.4341	.080
-.0841	.4491	.4283	.090
-.0812	.4490	.4234	.100
-.0613	.4347	.3914	.200
-.0479	.4154	.3703	.300
-.0376	.3966	.3529	.400
-.0295	.3791	.3377	.500
-.0229	.3629	.3240	.600
-.0174	.3481	.3116	.700
-.0129	.3344	.3002	.800
-.0091	.3218	.2896	.900
-.0059	.3101	.2798	1.000

جدول (١١-٦) قيم VIF لمعاملات الانحدار وقيم R^2 من أجل ثوابت الحياز مختلفة c. مثال شحوم الجسم بثلاثة متغيرات مستقلة.

R^2	$(VIF)_3$	$(VIF)_2$	$(VIF)_1$	c
.8014	104.61	564.34	708.84	0.00
.7943	19.28	100.27	125.73	.001
.7901	8.28	40.45	50.56	.002
.7878	4.86	21.84	27.18	.003
.7864	3.36	13.73	16.98	.004
.7854	2.58	9.48	11.64	.005
.7847	2.19	6.98	8.50	.006
.7842	1.82	5.38	6.50	.007
.7838	1.62	4.30	5.15	.008
.7834	1.48	3.54	4.19	.009
.7832	1.38	2.98	3.49	.010
.7818	1.01	1.08	1.10	.020
.7812	.92	.70	.63	.030
.7808	.88	.56	.45	.040
.7804	.85	.49	.37	.050
.7801	.83	.45	.32	.060
.7797	.81	.42	.30	.070
.7793	.79	.40	.28	.080
.7789	.78	.39	.26	.090
.7784	.76	.37	.25	.100
.7723	.63	.31	.21	.200
.7638	.54	.27	.18	.300
.7538	.46	.24	.17	.400
.7427	.40	.21	.15	.500
.7310	.35	.19	.14	.600
.7189	.31	.18	.13	.700
.7065	.28	.16	.12	.800
.6941	.25	.15	.11	.900
.6818	.23	.14	.11	1.000

تعليقات

١ـ المعادلات الناظمية (11.44) لمقدرات الحافة هي كما يلي:

$$\begin{aligned}
(1+c)b_1^R + r_{12}b_2^R + \ldots + r_{1,p-1}b_{p-1}^R &= r_{Y1} \\
r_{21}b_1^R + (1+c)b_2^R + \ldots + r_{2,p-1}b_{p-1}^R &= r_{Y2} \\
\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad &\quad \vdots \\
r_{p-1,1}b_1^R + r_{p-1,2}b_2^R + \ldots + (1+c)b_{p-1}^R &= r_{Yp-1}
\end{aligned} \tag{11.47}$$

حيث r_{ij} معامل الارتباط البسيط بين المتغير الـ i والمتغير الـ j من المتغيرات X، وهو r_{Yj} معامل الارتباط البسيط بين المتغير التابع Y والمتغير الـ j من المتغيرات X.

٢- تُعرَّف قيم الـ VIF لمعاملات انحدار الحافة b_k^R بصورة مشابهة لتلك الخاصة بمعاملات انحدار المربعات الدنيا المعتادة. أي أن قيم الـ VIF لـ b_k^R تقيس مدى كبر تباين b_k^R نسبة إلى ماكان سيكون عليه التباين لو أن المتغيرات المستقلة كانت غير مرتبطة. ويمكن تبيان أن قيم الـ VIF لمعاملات انحدار الحافة b_k^R هي العناصر القطرية للمصفوفة $(p-1)\times(p-1)$ التالية:

$$(\mathbf{r}_{XX}+c\mathbf{I})^{-1}\mathbf{r}_{XX}(\mathbf{r}_{XX}+c\mathbf{I})^{-1} \qquad (11.48)$$

٣- معامل التحديد المتعدد R^2، المعطى في (7.35) في حال المربعات الدنيا المعتادة على الشكل:

$$R^2 = 1-\frac{SSE}{SSTO} \qquad (11.49)$$

يمكن تعريفه بصورة مشابهة في حال انحدار الحافة. وعلى أي حال، فهناك تبسيط حاصل بسبب أن مجموع المربعات الكلي للمتغير التابع Y' الناتج عن تحويل الارتباط والمعطى في (8.41) هو:

$$SSTO_R = \sum(Y_i'-\bar{Y}')^2 = 1 \qquad (11.50)$$

والقيم التوفيقية في حالة انحدار الحافة هي:

$$\hat{Y}_i' = b_1^R X_{i1}' + \ldots + b_{p-1}^R X_{i,p-1}' \qquad (11.51)$$

حيث الـ X_{ik}' هي المتغيرات X بعد تحويلها وفقا لتحويل الارتباط (8.41b)، ومجموع مربعات الخطأ هو كالمعتاد:

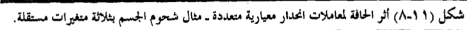
$$SSE_R = \sum(Y_i'-\hat{Y}_i')^2 \qquad (11.52)$$

حيث $\hat{Y}_i'$ معطى في (11.51) وعندئذ يصبح R^2 لانحدار الحافة:

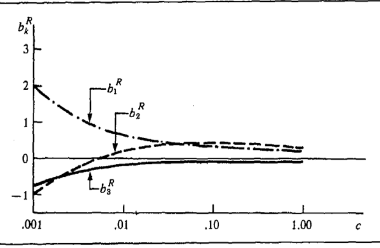
$$R_R^2 = 1-SSE_R \qquad (11.53)$$

٤- تميل تقديرات انحدار الحافة إلى أن تكون مستقرة، بمعنى أنها تتأثر في العادة، تأثرا بسيطا عند حصول تغيرات صغيرة في البيانات التي قام عليها الانحدار التوفيقي. وعلى العكس من ذلك، يمكن أن تكون تقديرات المربعات الدنيا المعتادة، تحت هذه الشروط، على درجة

عالية من عدم الاستقرار، وذلك عندما تكون المتغيرات المستقلة على درجة عالية من الخطية المتعددة. وأحيانا ستقدم دالة انحدار الحافة المقدّرة أيضا تقديرات جيدة لمتوسط استجابات مشاهدات جديدة أو تنبؤاتها وذلك من أجل مستويات للمتغيرات المستقلة خارج منطقة المشاهدات التي بُنيت عليها دالة الانحدار.



شكل (١١-٨) أثر الحافة لمعاملات انحدار معيارية متعددة ـ مثال شحوم الجسم بثلاثة متغيرات مستقلة.



وعلى العكس من ذلك، يمكن أن يكون أداء دالة الانحدار المقدّرة، القائمة على المربعات الدنيا المعتادة، أداء فقيرا تماما في ظروف كهذه. وبالطبع ينبغي دائما أن يتم أي تقدير أو تنبؤ يخرج بعيدا عن منطقة المشاهدات بحذر شديد.

٥ـ المحدودية الرئيسة لانحدار الحافة هو عدم إمكانية تطبيق طرق الاستقراء المعتادة، والخواص التوزيعية المضبوطة غير معروفة. والمحدودية الأخرى هي أن اختيار ثابت الانحياز c مسألة اجتهاد. وفي الوقت الذي طوّرت فيه طرق رسمية للقيام بهذا الاختيار إلا أن لهذه الطرق بدورها محدودياتها.

٦ـ عُمِّمت طرق انحدار الحافة بحيث تسمح بثوابت انحياز مختلفة لمعاملات انحدار مقدَّرة مختلفة.

تدابير علاجية أخرى

وهناك أساليب طُورت أيضا لعلاج مشاكل الخطية المتعددة، وتتضمن فيما تتضمن انحدار المركبات الرئيسة. حيث تكون المتغيرات المستقلة مركبات خطية في المتغيرات المستقلة الأصلية، وانحدار بايز حيث تُستوعب معلومات سابقة عن معاملات الانحدار في طريقة التقدير. ويمكن الحصول على مزيد من المعلومات عن هذه الأساليب وعن انحدار الحافة وانحدار الحافة المعمم أيضا من كتب متخصصة مثل المرجع [11.5].

(١١-٨) تدابير علاجية لتباينات خطأ غير متساوية - المربعات الدنيا المرجحة

شرحنا في الفصلين الرابع والسابع كيف يمكن لتحويلات المتغير التابع Y أن تكون مفيدة في تخفيض أو إلغاء عدم التساوي بين تباينات حدود الخطأ. والصعوبة في تحويلات Y أنها يمكن أن تبتدع علاقات انحدار غير مناسبة. وعند العثور على علاقة انحدار مناسبة، إلا أن تباينات حدود الخطأ غير متساوية، فإحدى البدائل هو المربعات الدنيا المرجحة.

المربعات الدنيا المرجحة

نبدأ شرحنا للمربعات الدنيا المرجحة بانحدار خطي بسيط. ومعيار المربعات الدنيا للانحدار الخطي البسيط، كما ورد في (2.8):

$$Q=\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1 X_i)^2$$

يعطي لكل مشاهدة Y الوزن نفسه. ومعيار المربعات الدنيا المرجحة لانحدار خطي بسيط يقدم أوزانا مختلفة:

$$Q_w=\sum_{i=1}^{n}w_i(Y_i-\beta_0-\beta_1 X_i)^2 \tag{11.54}$$

حيث w_i وزن المشاهدة Y وبجعل Q_w أصغر مايمكن بالنسبة لـ β_0 و β_1 نصل إلى المعادلات الناظمية:

$$\begin{aligned}\sum w_i Y_i &= b_0 \sum w_i + b_1 \sum w_i X_i \\ \sum w_i X_i Y_i &= b_0 \sum w_i X_i + b_1 \sum w_i X_i^2\end{aligned} \tag{11.55}$$

ويمكن حل هذه المعـادلات بدورها للوصول إلى مقـدرات المربعـات الدنيـا المرجحـة b_0 وb_1:

$$b_1 = \frac{\sum w_i X_i Y_i - \dfrac{\sum w_i X_i \sum w_i Y_i}{\sum w_i}}{\sum w_i X_i^2 - \dfrac{(\sum w_i X_i)^2}{\sum w_i}} \tag{11.56a}$$

$$b_0 = \frac{\sum w_i Y_i - b_1 \sum w_i X_i}{\sum w_i} \tag{11.56b}$$

لاحظ أنه إذا كانت جميع الأوزان متساوية بحيث تكون جميعها متطابقة وتساوي كميـة ثابتـة، فتُختَزل المعـادلات الناظميـة (11.55) للمربعـات الدنيـا المرجحـة إلى المعادلات في (2.9) الخاصة بالمربعات الدنيا غير المرجحة، وتُختزل مقـدِّرات المربعـات الدنيا المرجحة في (11.56) إلى مقدِّرات المربعات الدنيا غير المرجحة في (2.10).

ونعمم الآن المربعـات الدنيـا المرجحـة إلى انحـدار متعـدد ونقدمـه بصـورة أكـثر رسمية. وسنرمز بـ σ_i^2 لتباين حد الخطأ ε_i، ونعتبره مؤلفا من ثابت تناسب، نرمز له بـ σ^2، ومركبة w_i تختلف باختلاف حد الخطأ كما يلي:

$$\sigma_i^2 = \sigma^2\{\varepsilon_i\} = \sigma^2\{Y_i\} = \frac{\sigma^2}{w_i} \tag{11.57}$$

ويمكن لثابت التناسب أن يكون أي عدد موجب.

ومعيار المربعات الدنيا المرجحة لانحدار متعدد هو:

$$Q_w = \sum_{i=1}^{n} w_i (Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i,p-1})^2 \tag{11.58}$$

ونلاحظ من (11.57) أن الأوزان w_i متناسبة عكسيا مع التباينات σ_i^2. وهكذا تتلقـى مشاهدة Y_i تباينها كبير وزنا أقل من مشاهدة تباينها أصغر. وكلما كـانت أكـثر دقـة (أي كلما كان σ_i^2 أصغر) كلما كانت المعلومات الـتي تقدمهـا Y_i عـن$E\{Y_i\}$ أكـثر، وكلما استحقت وزنا أكبر عند توفيق دالة انحدار.

لتكن المصفوفة $\mathbf{W}$ مصفوفة قطرية تتضمن الأوزان w_i:

$$\underset{n\times n}{\mathbf{W}}=\begin{bmatrix} w_1 & & & & 0 \\ & w_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & w_n \end{bmatrix} \qquad (11.59)$$

فيمكن عندئذ التعبير عن المعادلات الناظمية للمربعات الدنيا المرجحة كما يلي:

$$(\mathbf{X'WX})=\mathbf{X'WY} \qquad (11.60)$$

ومقدرات المربعات الدنيا المرجحة لمعاملات الانحدار هي:

$$\underset{p\times 1}{\mathbf{b}}=(\mathbf{X'WX})^{-1}\mathbf{X'WY} \qquad (11.61)$$

ومصفوفة التباين ـ التغاير للمعاملات المقدرة لانحدار المربعات الدنيا المرجحة هي:

$$\underset{p\times p}{\sigma^2\{\mathbf{b}\}}=\sigma^2(\mathbf{X'WX})^{-1} \qquad (11.62)$$

حيث σ^2 ثابت التناسب في (11.57). ومصفوفة التباين ــ التغاير المقدرة لمعاملات الانحدار هي:

$$\underset{p\times p}{\mathbf{s}^2\{\mathbf{b}\}}=MSE_w(\mathbf{X'WX})^{-1} \qquad (11.63)$$

حيث يستند MSE_w إلى المربعات المرجحة للانحرافات:

$$MSE_w=\frac{\sum w_i(Y_i-\hat{Y}_i)^2}{n-p} \qquad (11.63a)$$

وهكذا يكون MSE_w هنا مقدَّرا لثابت التناسب σ^2.

الأوزان عندما تكون σ_i^2 مجهولة

إذا كانت التباينات σ_i^2 معروفة تماما أو مقربة إلى ثابت تناسب، فسيكون استخدام المربعات الدنيا المرجحة بأوزان w_i أمرا سهلا. ومن سوء الحظ فمن النادر معرفة التباينات σ_i^2 مما يضطرنا إلى استخدام تقديرات للتباينات. ويمكن الحصول على هذه التباينات المقدَّرة بطرق متنوعة. ونناقش هنا طريقتين للحصول على تقديرات التباينات σ_i^2.

١- تتغير تباينات حد الخطأ أحيانا مع تغير المستوى لمتغير مستقل في نموذج الانحدار، وذلك بطريقة منظمة. وعلى سبيل المثال، يمكن أن تكون العلاقة، في حالة انحدار خطي بسيط إحدى العلاقات التالية:

$$\sigma_i^2 = \sigma^2 X_i \tag{11.64a}$$

$$\sigma_i^2 = \sigma^2 X_i^2 \tag{11.64b}$$

$$\sigma_i^2 = \sigma^2 \sqrt{X_i} \tag{11.64c}$$

وهنا نجد σ^2 من جديد كثابت تناسب.

وإذا كان تحليل الرواسب يقترح، مثلا، نزوع σ_i^2 إلى التغير مباشرة مع X_i، فعندئذ يمكن استخدام الأوزان:

$$w_i = \frac{1}{X_i}$$

وبصورة مماثلة يمكن أن تكون أوزان الحالتين (11.64b) و (11.64c)، $w_i = 1/X_i^2$ و $w_i = 1/\sqrt{X_i}$ على الترتيب.

٢- وعندما تتغير تباينات حد الخطأ مع تغير المستوى لمتغير مستقل، ولكن بصورة غير منتظمة، يمكن تجميع المشاهدات في عدد صغير من المجموعات وفقا لمستوى المتغير المستقل، وتُحسب تباينات الرواسب لكل مجموعة. وعندئذ تتلقى كل مشاهدة Y من مجموعة وزنا هو مقلوب التباين المقدَّر لتلك المجموعة. ويمكن استخدام هذا الإجراء نفسه عندما يتغير تباين حد الخطأ في انحدار متعدد مع مستوى القيمة التوفيقية $\hat{Y}$. وهنا تُجمع المشاهدات وفقا لقيمها التوفيقية.

ويمكن أن تكون هذه الطرق التقريبية للترجيح مفيدة جدا عندما يشير تحليل الرواسب إلى فروق مهمة في تباينات حدود الخطأ. وعندما تكون الفروق صغيرة أو متواضعة فسوف لاتكون المربعات الدنيا المرجحة مفيدة، على وجه الخصوص، مع مثل هذه الطرق التقريبية.

جدول (١١-٧) بيانات ضغط الدم الانبساطي

الشخص i	العمر X_i	ضغط الدم الانبساطي Y_i	الشخص i	العمر X_i	ضغط الدم الانبساطي Y_i
1	27	73	28	49	101
2	21	66	29	40	70
3	22	63	30	42	72
4	26	79	31	43	80
5	25	68	32	46	83
6	28	67	33	43	75
7	24	75	34	49	80
8	25	71	35	40	90
9	23	70	36	48	70
10	20	65	37	42	85
11	29	79	38	44	71
12	24	72	39	46	80
13	20	70	40	47	96
14	38	91	41	45	92
15	32	76	42	55	76
16	33	69	43	54	71
17	31	66	44	57	99
18	34	73	45	52	86
19	37	78	46	53	79
20	38	87	47	56	92
21	33	76	48	52	85
22	35	79	49	57	109
23	30	73	50	50	71
24	37	68	51	59	90
25	31	80	52	50	91
26	39	75	53	52	100
27	46	89	54	58	80

مثال.

تهتم باحثة صحية بدراسة العلاقة بين ضغط الدم الانبساطي والعمر عند النساء البالغات اللواتي يتمتعن بصحة جيدة وتتراوح أعمارهن بين 20 و 60 عاما، وقد جمعت بيانات إحصائية عن 54 امرأة. والبيانات مقدمة في الجدول (١١–٧). ويقترح

مخطط الانتشار في الشكل (١١-٩)أ، بقوة علاقة خطية بين ضغط الدم الانبساطي والعمر ولكنه يشير إلى زيادة تباين حد الخطأ مع العمر. وقامت الباحثة بتوفيق دالة انحدارخطية مستخدمة المربعات الدنيا غير المرجحة وذلك للقيام ببعض التحليلات الأولية للرواسب، ورسم الرواسب في مقابل X، المقدم في الشكل (١١-٩)ب، يؤكد عدم ثبات تباين الخطأ.

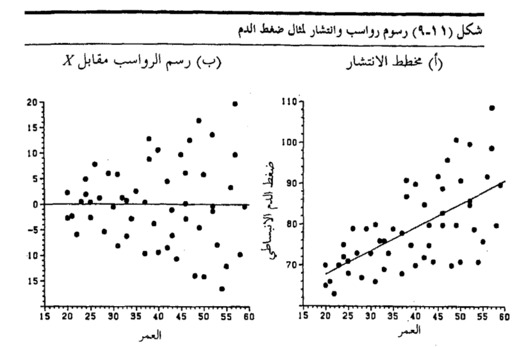
ولاستكشاف ما إذا كان لتباين حد الخطأ علاقة بسيطة بالعمر، قسّمت الباحثة المشاهدات إلى أربع مجموعات عُمرية لها تقريبا الحجم نفسه. وعندئذ تمَّ، من أجل كل مجموعة، حساب تباين العينة للرواسب الناتجة عن انحدار مربعات دنيا غير مرجحة. والمجموعات العمرية الأربع وعدد المشاهدات في كل منها مقدمة في العمودين الأول والثاني من الجدول (١١-٨)، وفي العمود الثالث قُدّمت التباينات المقدَّرة للخطأ. وقد اعتبرت الباحثة الفروق في التباينات كبيرة، مما يدعو إلى استخدام المربعات الدنيا المرجحة، وقد تفحصت التباينات لرؤية ما إذا كانت تتبع أيا من القواعد المذكورة في (11.64) مستخدمة النقطة المتوسطة من مدى الأعمار لكل مجموعة كقيمة لـ X وقد حصلت على النتائج التالية:

الزمرة j	X_j	s_j^2/X_j	s_j^2/X_j^2	$s_j^2/\sqrt{X_j}$
1	25	.71	.028	3.5
2	35	1.20	.034	7.1
3	45	1.95	.043	13.1
4	55	2.26	.041	16.7

جدول (١١-٨) تباينات الخطأ المقدّرة والأوزان للمجموعات العمرية - مثال ضغط الدم.

الزمرة j	العمر X_j	حجم العينة	تباين الخطأ المقدّر s_j^2	الوزن المقدّر $w_j = 1/s_j^2$
1	20 - under 30	13	17.74260	.0563615
2	30 - under 40	13	42.13678	.0237322
3	40 - under 50	15	87.93657	.0113718
4	50 - under 60	13	124.14565	.0080551

شكل (١١-٩) رسوم رواسب وانتشار لمثال ضغط الدم



ولم تعتبر أيا من هذه العلاقات مستقرة بصورة كافية، وقررت بالتالي استخدام مقلوب التباينات كأوزان لكل مشاهدة في المجموعة. وهذه الأوزان مبينة في العمود الرابع من الجدول (١١-٨).

وقد أنتج برنامج حاسب لتحليل انحدار مرجّح خط الانحدار التوفيقي التالي:

$$\hat{Y} = 56.15693 + 0.580031X \qquad (11.65)$$

وخط الانحدار التوفيقي مبيّن في الشكل (١١-٩)أ ويبدو أنه توفيق جيد إلى حد ما للبيانات.

وخط الانحدار التوفيقي المقدَّر للبيانات نفسها مستخدمين المربعات الدنيا غير المرجّحة هو:

$$\hat{Y} = 56.08962 + 0.589583X \qquad (11.66)$$

وهو يختلف إلى حد ما عن خط المربعات الدنيا المرجحة في (11.65)، كما سيكون عليه الحال بصورة عامة، ولكن الفروق هنا ليست كبيرة.

وبينما تكون التقديرات، التي نحصل عليها بطريقة المربعات الدنيا غير المرجحة، تقديرات غير منحازة، حتى عندما تكون تباينات الخطأ غير متساوية، شأنها في ذلك شأن التقديرات الناتجة عن المربعات الدنيا المرجحة، فإن تقديرات المربعات الدنيا غير المرجحة تخضع لتغير معاينة أكبر. وفي مثالنا، نجد الانحرافات المعيارية المقدرة لمعاملات الانحدار في الطريقتين كما يلي:

مربعات دنيا مرجحة	مربعات دنيا غير مرجحة
$s\{b_0\} = 2.7908$	$s\{b_0\} = 3.9937$
$s\{b_1\} = 0.08401$	$s\{b_1\} = 0.09695$

تعليقات

١- يدعى شرط عدم ثبات تباين الخطأ فوق جميع المشاهدات "عدم تجانس" خلافا لشرط تساوي تباينات الخطأ الذي يدعى "تجانس".

٢- عندما يسود عدم التجانس مع تحقق الشروط الأخرى لنموذج الانحدار (7.18)، تبقى معاملات الانحدار المقدَّرة التي نحصل عليها بطرق المربعات الدنيا العادية غير منحازة ومتّسقة، إلا أنها لاتعود مقدَّرات غير منحازة ذات تباين أصغري، كما هو موضح في المثال السابق.

٣- خاصية عدم التجانس هي خاصية متأصلة عندما تتبع الاستجابة في تحليل الانحدار توزيعا يكون التباين فيه على صلة دالّية بالمتوسط. (وفي معظم هذه الحالات نواجه حيدانا مهما لـ Y عن الناظمية) لنعتبر، في هذا السياق تحليل انحدار حيث X سرعة آلة تضع غلافا بلاستيكيا لكَبْل وY عدد العيوب في التغليف لكل ألف قدم من الكبل إذا كانت Y تتوزع وفق بواسون بمتوسط يزداد بازدياد X، فلا يمكن أن يكون لتوزيعات Y تباين ثابت عند مستويات X نظرا لأن تباين متغير بواسون يساوي متوسطه، والمتوسط يزداد مع X .

٤- للمشاهدات المكررة فائدة جمة في الحصول على معلومات حول أية نمطية في تباينات الخطأ. إلا أن المشاهدات المكررة لاتتوافر، في الغالب، وتبرز الحاجة عندئذ إلى

استخدام تجميعات لمشاهدات متساوية الحجم تقريبا، كما في مثال ضغط الدم، وذلك للحصول على معلومات عن تباينات الخطأ.

٥- يمكن استخدام طريقة المربعات الدنيا المرجحة المكررة لتحسين تقديرات المربعات الدنيا المرجحة. وتنطوي هذه الطريقة على تقدير مبدئي للأوزان من البيانات، ثم الحصول على دالة الانحدار التوفيقية والرواسب بطريقة المربعات الدنيا المرجحة. وباستخدام الرواسب من هذه المرحلة الأولى، نعيد تقدير الأوزان w_i لنحصل على توفيق مربعات دنيا مرجحة جديد. وتستمر العملية حتى تصبح التغيرات الحاصلة في دالة الانحدار التوفيقية غير ذات جدوى. وفي الغالب، يكون تكرار واحد أو تكراران كافياً.

٦- يمكن الحصول على تقديرات مربعات دنيا مرجحة باستخدام مربعات دنيا غير مرجحة على متغيرات حولناها بصورة مناسبة. وعلى سبيل المثال، لنعتبر انحدارا خطيا بسيطا تكون σ_i^2 فيه متناسبة مع X_i^2، بحيث تكون الأوزان $w_i = 1/X_i^2$ ويكون معيار المربعات الدنيا المرجحة (11.54) عندئذ كما يلي:

$$Q_w = \sum \frac{1}{X_i^2}(Y_i - \beta_0 - \beta_1 X_i)^2 = \sum\left(\frac{Y_i}{X_i} - \frac{\beta_0}{X_i} - \beta_1\right)^2 \tag{11.67}$$

ويمكن التعبير عن هذا المعيار على الشكل:

$$Q_w = \sum (Y_i' - \beta_0' - \beta_1' X_i')^2 \tag{11.67a}$$

حيث:

$$Y_i' = \frac{Y_i}{X_i} \qquad \beta_0' = \beta_1 \qquad X_i' = \frac{1}{X_i} \qquad \beta_1' = \beta_0$$

ونلاحظ أن (11.67a) تتخذ شكل معيار المربعات الدنيا غير المرجحة (2.8). وبالتالي يمكن تطبيق المربعات الدنيا العادية على المشاهدات المحوّلة Y_i' و X_i' لتنتج التقديرات نفسها، التي تنتجها المربعات الدنيا المرجحة، مطبقة على المشاهدات الأصلية. ويمكن تبيان أن تباين الخطأ للمتغير المحوّل Y' ثابت.

٧- المربعات الدنيا المرجحة هي حالة خاصة من مربعات دنيا معممة حيث يمكن لحدود الخطأ، لا أن يكون لها تباينات مختلفة فحسب، ولكن يمكن لأزواج من حـدود الخطأ أن تكون مرتبطة أيضا.

مراجع ورد ذكرها

[11.1] Atkinson, A. C. *Plots, Transformations, and Regression.* Oxford : Clarendon Press, 1985.

[11.2] Mansfield, E. R., and conerly, M. D."Diagnostic Value of Residual and Partial Residual Plots." *The American Statistician* 41 (1987), 107 - 16.

[11.3] Kennedy, W. J., Jr., and Gentle, J. E. *Statistical Computing.* New York : Marcel Dekker, 1980.

[11.4] Hogg, R. V. "Statistical Robustness: One View of its Use in Applications Today." *The American Statistician* 33 (1979),108-15.

[11.5] Belsley, D. A; Kuh, E. and Welsch, R. E. *Regression Diagnostics : Identifying Influential Data and Sources of Collinearity.* New York : John Wiley & Sons, 1980.

مسائل

(١١-١) سأل طالب: "لماذا يكون القيام بتحريات تشـخيصية لعملية توفيق ضروريا عندما يكون R^2 كبيرا ؟" علّق.

(١١-٢) صرح باحث: "أحد الميزات الطيبة لرسومات انحدار جزئـي هـي أنهـا مفيـدة للغاية في التحقق من صلاحية نموذج حتـى عندمـا تكـون المتغـيرات المسـتقلة مرتبطة ارتباطا عاليا". علّق.

(١١-٣) اقترح طالب: "إذا اكتشفت وجود مشاهدات قاصية واسعة النفـوذ في مجموعـة بيانات، فاحذف هذه المشاهدات ببساطة من مجموعة البيانات". علّق.

(١١-٤) صف عدة طرق غير رسمية مما يمكن أن يكـون مفيـدا في التحقـق مـن وجـود خطية متعددة بين المتغيرات المستقلة في نموذج انحدار متعدد.

(١١-٥) بالإشارة إلى مسألة **تفضيل صنف** (٧-٨)أ.

أ - قم بإعداد رسم انحدار جزئي لكل من المتغيرات المستقلة.

ب ـ هل تقترح رسوماتك في الجزء (أ) أن علاقات الانحدار في دالة الانحدار التوفيقية في المسألة (٧ـ٨)أ غير مناسبة لأي من المتغيرات المستقلة؟ اشرح.

جـ ـ أوجد دالة الانحدار التوفيقية في المسألة (٧ـ٨)أ بأن تحدر أولا كلا من Y و X_2 على X_1، ثم احدر الرواسب بطريقة مناسبة.

(١١ـ٦) بالإشارة إلى **شحنة الكيماويات** في المسألة (٧ـ١٢)جـ.

أ ـ قم بإعداد رسم انحدار جزئي لكل من المتغيرات المستقلة.

ب ـ هل تقترح رسوماتك في الجزء (أ) أن علاقات الانحدار في دالة الانحدار التوفيقية في المسألة (٧ـ١٢)جـ غير مناسبة لأي من المتغيرات المستقلة؟ اشرح.

جـ ـ أوجد دالة الانحدار التوفيقية في المسألة (٧ـ١٢)جـ بأن تحدر أولا كلا من Y و X_2 على X_1 ثم احدر الرواسب بطريقة مناسبة.

(١١ـ٧) بالإشارة إلى **ارتياح مريض** مسألة (٧ـ١٧)ب.

أ ـ قم بإعداد رسم انحدار جزئي لكل من المتغيرات المستقلة.

ب ـ هل تقترح رسوماتك في الجزء (أ) أن علاقات الانحدار في دالة الانحدار التوفيقية في المسألة (٧ـ١٧)ب غير مناسبة لأي من المتغيرات المستقلة؟ اشرح.

(١١ـ٨) بالإشارة إلى **رواتب المختصين في الرياضيات** مسألة (٧ـ٢٠)ب.

أ ـ قم بإعداد رسم انحدار جزئي لكل من المتغيرات المستقلة.

ب ـ هل تقترح رسوماتك في الجزء (أ) أن علاقات الانحدار في دالة الانحدار التوفيقية في المسألة (٧ـ٢٠) غير مناسبة لأي من المتغيرات المستقلة؟ اشرح.

(١١ـ٩) بالإشارة إلى **تفضيل صنف** مسألة (٧ـ٨)، العناصر القطرية لمصفوفة القبّعة هي:

$$h_{55} = h_{66} = h_{77} = h_{88} = h_{99} = h_{10,10} = h_{11,11} = h_{12,12} = 0.137$$

و $$h_{11} = h_{22} = h_{33} = h_{44} = h_{13,13} = h_{14,14} = h_{15,15} = h_{16,16} = 0.237$$

أ - اشرح سبب النمطية في العناصر القطرية لمصفوفة القبعة.

ب ـ وفقا لقاعدة إصبع الإبهام المعروضة في هذا الفصل، هل أي من المشاهدات قاصية بالنسبة لقيمها وفقا لـ X ؟.

جـ - أوجد رواسب الحذف المعيّرة تقديرا وحدد أية مشاهدات قاصية بالنسبة لِـ Y.

د - تبدو المشاهدة 14 مشاهدة قاصية بالنسبة لِـ Y . أوجد قيم ***DFFITS***، ***DFBETAS*** ومسافة كوك لهذه المشاهدة لتقويم نفوذها. ماذا تستنتج؟

هـ - احسب للقيم التوفيقية متوسط مطلق الفروق النسبية المئوية مع المشاهدة 14 وبدونها. إلام يشير هذا المقياس بالنسبة لنفوذ المشاهدة 14؟

و - احسب مسافة كوك D_i لكل مشاهدة. هل هناك أية مشاهدات ذات نفوذ وفقا لهذا المقياس؟.

(١١-١٠) بالإشارة إلى **شحنة الكيماويات** مسألة (٧-١٢)، كانت العناصر القطرية لمصفوفة القبعة كما يلي:

i	1	2	3	4	5	6	7	8	9	10
h_{ii}	.091	.194	.131	.268	.149	.141	.429	.067	.135	.165

i	11	12	13	14	15	16	17	18	19	20
h_{ii}	.179	.051	:110	.156	.095	.128	.097	.230	.112	.073

أ - حدد أية مشاهدات قاصية في X مستخدما قاعدة إصبع الإبهام التي وردت في هذا الفصل.

ب - أوجد رواسب الحذف المعيّرة وحدد أية مشاهدات قاصية في Y.

جـ - يبدو أن المشاهدة 7 قاصية في X والمشاهدة 12 قاصية في Y. احسب قيم ***DFFITS***، ***DFBETAS***، ومسافة كوك لكل من هاتين المشاهدتين لتثمين تأثيرهما. ماذا تستنتج؟

د - احسب متوسط مطلق الفرق النسبي المئوي للقيم التوفيقية مع المشاهدة 7 وبدونها، ومع المشاهدة 12 وبدونها. إلى ماذا يشير هذا المقياس فيما يتعلق بتأثير كل من المشاهدتين ؟

هـ ـ احسب مسافة كوك D_i لكل مشاهدة. هل هناك أية مشـاهدات مؤثـرة وفقا لهذا المقياس؟.

(١١-١١) بالإشـارة إلى ارتيـاح **المريـض** مسـألة (٧ـ١٧). كـانت العنـاصر القطريـة لمصفوفة القبعة كما يلي:

10	9	8	7	6	5	4	3	2	1	i
.137	.339	.174	.060	.319	.204	.235	.070	.193	.134	h_{ii}

20	19	18	17	16	15	14	13	12	11	i
.231	.078	.143	.209	.104	.313	.072	.230	.057	.245	h_{ii}

23	22	21	i
.059	.238	.158	h_{ii}

أ ـ حدد أية مشاهدات قاصية في X .

ب ـ أوجد رواسب الحذف المعيّرة تقديرا وحدّد أية مشاهدات قاصية في Y.

جـ ـ تبدو المشاهدة 14 بأنها واقعة على حدود المشـاهدات القاصيـة في Y. احسب قيم ***DFFITS***، ***DFBETAS***، ومسافة كـوك لهـذه المشـاهدة وذلك لتقويم نفوذها. ماذا تستنتج؟.

د ـ احسب متوسط مطلق الفرق النسبي المئوي للقيم التوفيقية مع المشاهدة 14 وبدونها إلام يشير هذا المقياس فيما يتعلق بنفوذ المشاهدة 14 ؟

هـ ـ احسب مسافة كوك لكل مشاهدة. هـل هنـاك أيـة مشـاهدات نـافذة وفقا لهذا المقياس؟.

(١١-١٢) بالعودة إلى **رواتب المختصين في الرياضيات** مسألة (٧ـ٢٠). **فإن** العنـاصر القطرية لمصفوفة القبعة هي كما يلي:

8	7	6	5	4	3	2	1	i
.179	.115	.146	.214	.071	.132	.059	.184	h_{ii}

16	15	14	13	12	11	10	9	i
.151	.186	.098	.320	.128	.083	.288	.241	h_{ii}

24	23	22	21	20	19	18	17	i
.110	.225	.136	.118	.206	.198	.146	.267	h_{ii}

أ ـ حدد أية مشاهدات قاصية في X.

ب ـ أوجد رواسب الحذف المعيّرة تقديرا، وحدد أية مشاهدات قاصية في Y.

جـ ـ تبدو المشاهدة 19 بأنها واقعة على حدود المشاهدات القاصية في Y. احسب قيم *DFFITS* ، *DFBETAS*، ومسافة كوك لهذه المشاهدة وذلك لتقويم نفوذها. ماذا تستنتج؟.

د ـ احسب متوسط مطلق الفرق النسبي المئوي للقيم التوفيقية مع المشاهدة 19 وبدونها. إلام يشير هذا المقياس فيما يتعلق بنفوذ المشاهدة 19؟

هـ ـ احسب D_i مسافة كوك لكل مشاهدة. هل هناك أية مشاهدات نافذة وفقا لهذا المقياس؟.

(١١-١٣) **مبيعات مواد التجميل**. حصل مساعد في مكتب مبيعات منطقة لشركة مواد تجميل وطنية على البيانات المعروضة أدناه، والمتعلقة بنفقات الدعاية والمبيعات في العام الماضي في الدوائر الأربع عشرة في المنطقة. ويرمز X_1 لنفقات العرض في صالونات التجميل ومتاجر المنوعات (بآلاف الدولارات) بينما تمثل X_2 وX_3 على الترتيب، النفقات المقابلة للدعاية في وسائل الإعلام المحلية أوالحصة المخصصة من نفقات الدعاية في وسائل الإعلام القومية. وترمز Y للمبيعات (بآلاف الحالات)، وقد طُلب من المساعد تقدير الزيادة في المبيعات المتوقعة عندما يزيد X_1 ألف دولار مع بقاء X_2 وX_3 ثابتين، كما أعطي التوجيهات باستخدام نموذج الانحدار المتعدد العادي بحدود خطية في المتغيرات المستقلة وحدود خطأ ناظمية مستقلة.

7	6	5	4	3	2	1	i
4.8	7.2	2.9	2.1	3.0	6.5	4.2	X_{i1}
5.0	7.0	3.0	2.0	3.5	6.5	4.0	X_{i2}
4.5	3.0	4.0	3.0	4.0	5.0	3.0	X_{i3}
8.56	12.18	7.84	5.62	9.73	14.70	8.26	Y_i

14	13	12	11	10	9	8	i
3.0	2.2	5.5	6.2	3.1	2.6	4.3	X_{i1}
2.8	2.0	5.5	6.0	3.0	2.5	4.0	X_{i2}
3.0	4.0	5.0	4.5	4.0	5.0	5.0	X_{i3}
6.74	7.15	10.46	12.51	8.90	7.56	10.77	Y_i

أ ـ اعرض نموذج الانحدار الذي سيستخدم في توفيق البيانات.

ب ـ اختبر ما إذا كانت هناك علاقة انحدار بين المبيعات والمتغيرات المستقلة الثلاثة. استخدم مستوى معنوية 0.05 اعرض البديلين وقاعدة القرار والنتيجة.

جـ ـ اختبر لكل من معاملات الانحدار $\beta_k(k = 1, 2, 3)$ على حدة ما إذا كان $\beta_k = 0$. استخدم مستوى معنوية 0.05 في كل اختبار. هل تتفق نتائج هذه الاختبارات مع نتيجة الاختبار في الجزء ب؟.

د ـ أوجد مصفوفة الارتباط للمتغيرات X.

هـ ـ ماذا تقترح النتائج في ب، جـ ود حول تلاؤم البيانات مع هدف البحث.

(١١-١٤) بالعودة إلى **مبيعات مواد التجميل**، المسألة (١١-١٣).

أ – تحقق من أن عامل تضخم التباين للمتغير X_1 هو $(VIF)_1 = 66.29$ وأن عاملي تضخم التباين الآخرين هما $(VIF)_2 = 66.99$ و$(VIF)_3 = 1.09$. ماذا تقترح هذه القيم بالنسبة لتأثيرات الخطية المتعددة هنا؟

ب ـ قرر المساعد أخيرا شطب المتغيرين X_2 وX_3 من النموذج بغية "جلاء الصورة". قم بتوفيق النموذج المحسن للمساعد. هل المساعد الآن في وضع أفضل لإنجاز هدف البحث؟

جـ ـ لماذا لاتكون التجربة هنا أكثر فعالية في تقديم بيانات مناسبة لمواجهة هدف البحث؟كيف تصمم تجربة كهذه؟ ماهو نموذج الانحدار الذي ستستخدمه؟.

(١١-١٥) بالعودة إلى **ارتياح المريض** مسألة (٧-١٧).

أ ـ أوجد مصفوفة الارتباط للمتغيرات X. ماذا توضح هذه المصفوفة حول الصلة الخطية بين أزواج المتغيرات المستقلة؟

ب ـ عوامل تضخم التباين هي $(VIF)_1 = 1.35$، $(VIF)_2 = 2.76$ و$(VIF)_3 = 2.87$

ماذا تقترح هذه النتائج حول الخطية المتعددة هنا؟ هل هذه النتائج أنجح في الكشف عن الخطية المتعددة من النتائج في الجزء (أ)؟.

(١١-١٦) بالعودة إلى **تفضيل صنف** مسألة (٧-٨).

أ ـ أوجد مصفوفة الارتباط للمتغيرات X. ماذا تبيّن هذه المصفوفة حول الصلة الخطية بين أزواج المتغيرات المستقلة؟.

ب ـ أوجد عاملي تضخم التباين، لماذا تجد كلا منهما مساويا للواحد؟

(١١-١٧) بالعودة إلى رواتب **المختصين في الرياضيات**، مسألة (٧-٢٠).

أ ـ أوجد مصفوفة الارتباط بين المتغيرات X . ماذا تبيّن هذه المصفوفة حول الصلة الخطية بين المتغيرات المستقله؟

ب ـ أوجد عوامل تضخم التباين. هل تشير إلى وجود مشكلة خطية متعددة خطرة هنا؟

(١١-١٨) بالعودة إلى مبيعات **مواد التجميل**، مسألة (١١-١٣). نعطي أدناه معاملات انحدار الحافة المعيارية المقدَّرة، عوامل تضخم التباين، وR^2 من أجل مختارات من قيم ثابت الانحياز.

c	.000	.005	.01	.02	.03	.04	.05	.06
b_1^R	.273	.327	.349	.368	.376	.380	.382	.383
b_2^R	.549	.494	.470	.447	.435	.427	.422	.417
b_3^R	.260	.260	.260	.259	.257	.256	.254	.253
$(VIF)_1$	66.29	24.11	12.45	5.20	2.92	1.91	1.38	1.07
$(VIF)_2$	66.99	24.36	12.57	5.25	2.94	1.92	1.39	.107
$(VIF)_3$	1.09	1.06	1.04	1.01	.99	.97	.95	.93
R^2	.8402	.8401	.8401	.8401	.8398	.8397	.8395	.8393

أ ـ ارسم أثر الحافة من أجل قيم المعطاة، هل تظهر معاملات انحدار الحافة تغيرات مرموقة في جوار $c = 0$؟

ب ـ اقترح قيمة معقولة لثابت الانحياز c، تستند إلى أثر الحافة، قيم الـ VIF وR^2.

جـ ـ حوّل معاملات الانحدار المعيارية المقدَّرة التي اخترتها في الجزء (ب) عـائدا إلى المتغيرات الأصلية، وأوجد القيم التوفيقية للمشاهدات الأربع عشـرة. ماهي درجة الشبه بين هذه القيم التوفيقية وتلك التي حصلـت عليهـا عند توفيق المربعات الدنيا العادية في المسألة (١١-٣)أ؟.

(١١-١٩) بالعودة إلى **شحنة الكيماويات**، المسألة (٧-١٢). نعطي أدنـاه معـاملات انحدار الحافة المعيارية المقدّرة، وعوامل تضخم التباين، وR^2 لمختـارات مـن ثوابت الانحياز c.

c	.000	.005	.01	.05	.07	.09	.10	.20
b_1^R	.451	.453	.455	.460	.460	.459	.458	.444
b_2^R	.561	.556	.552	.526	.517	.508	.504	.473
b_3^R	7.03	6.20	5.51	2.65	2.03	1.61	1.46	.71
$(VIF)_1=(VIF)_2$	.9869	.9869	.9869	.9862	.9856	.9852	.9844	.9780

أ ـ ارسم أثر الحافـة للقيـم المعطـاة. هـل تظهـر معـاملات انحـدار الحافـة تغيرات مرموقة في c في جوار $c = 0$؟

ب ـ لماذا يتساوى $(VIF)_1$ و $(VIF)_2$ هنا؟

جـ ـ اقترح قيمة معقولة لثابت الانحيـاز c مسـتندا إلى أثـر الحافـة في الجـزء (أ)، قيم الـ VIF و R^2 .

د ـ حوّل معاملات الانحدار المعيارية المقدَّرة التي اخترتهـا في الجـزء (جـ) عـائدا إلى المتغـيرات الأصليـة، وأوجـد القيـم التوفيقيـة للمشـاهدات العشرين. ماهي درجة الشبه بين هذه القيم التوفيقية وتلك الـتي حصلـت عليها عند توفيق المربعات الدنيا العادية في المسألة (٧-١٢)جـ.؟

(١١-٢٠) **سرعة آلة**. من المعروف أن عدد القطـع المعيبـة Y الـتي تنتجهـا آلـة يرتبـط خطيا بعيار السرعة X للآلة. وقـد جُمعـت البيانـات أدنـاه مـن سـجلات حديثة لضبط الجودة.

i	1	2	3	4	5	6	7	8	9	10	11	12
X_i	200	400	300	400	200	300	300	400	200	400	200	300
Y_i	28	75	37	53	22	58	40	96	46	52	30	69

أ ـ أوجد نموذج الانحدار التوفيقي (3.1) ودالة الانحدار المقدَّرة، وارسم الرواسب في مقابل X، ماذا يبين رسم الرواسب؟.

ب ـ احسب تباين العينة s^2 للرواسب، وذلك من أجل كل من سرعات الآلة الثلاث: $X = 200, 300, 400$ ماذا تقترح تباينات العينة الثلاثة حول ما إذا كانت تباينات حد الخطأ عند المستويات الثلاثة لـ X متساوية أم لا؟.

جـ ـ احسب s^2/X ، s^2/X^2 و $s^2/\sqrt{X}$ لكل من السرعات الثلاث للآلة. هل تبدو أي من هذه العلاقات مستقرة؟.

د ـ مستخدما الأوزان $w_i = 1/X_i^2$ أوجد تقديرات المربعات الدنيا المرجحة لِ β_0 و β_1. هل هذه التقديرات مشابهة لتلك التي وجدناها باستخدام المربعات الدنيا العادية في الجزء (أ)؟.

هـ ـ قارن الانحرافات المعيارية المقدَّرة لتقديرات المربعات الدنيا المرجحة b_0 وb_1 في الجزء (د) بتلك الخاصة بتقديرات المربعات الدنيا العادية في الجزء (أ)، ماذا تجد؟

(١١-٢١) **التعلم بمساعدة الحاسب.** فيما يلي بيانات من دراسة حول التعلم بمساعدة الحاسب لـ 12 طالبا، وهي تبين عدد الاستجابات الكلي في إتمام درس X وكلفة زمن الحاسب (بالسنتات).

12	11	10	9	8	7	6	5	4	3	2	1	i
11	18	12	19	13	10	17	14	10	22	14	16	X_i
51	81	57	88	63	55	70	62	50	85	70	77	Y_i

أ ـ قم بتوفيق دالة انحدار خطية مستخدما المربعات الدنيا العادية، أوجد الرواسب، وارسم الرواسب في مقابل X. ماذا تقترح رسوم الراسب؟

ب ـ جمّع المشاهدات التي تتضمن $X = 10, 11, 12$ في مجموعة واحدة، تتضمن $X = 13, 14, 16$ في مجموعة أخرى، وتتضمن $X = 17, 18, 19, 22$ في مجموعة ثالثة. ولكل من هذه المجموعات احسب تباين العينة للرواسب. هل يبدو أن تباينات الخطأ متساوية؟.

جـ - لكل مـن المجموعـات الثـلاث احسـب s^2 / X، s^2 / X^2 و $s^2 / \sqrt{X}$ مستخدما، كقيمة لـِ X، النقطة المتوسـطة لقيـم X في المجموعـة. هـل تبدو أي من هذه العلاقات مستقرة؟

د - مسـتخدما الأوزان $w_i = 1/X_i^2$ أوجـد تقديـرات المربعـات الدنيـا المرجحة لـِ β_0 و β_1 . هل هذه التقديرات مشابهة لتلك التي حصلت عليها في الجزء (أ) بطريقة المربعات الدنيا العادية؟

هـ - قارن الانحرافات المعيارية المقدّرة لتقديـرات المربعات الدنيـا المرجحـة b_0 و b_1 في الجزء (د) بتلك الخاصة بتقديرات المربعات الدنيـا العادية في الجزء (أ). ماذا تجد؟.

(١١-٢٢) بالإشارة إلى مثال ضغط الدم في الجدول (١١-٧). استنتج محلل قام بمراجعة نتائج الباحث على الصفحة (٥٤٦)والمتعلقة بإمكانية وجود علاقة بسيطة بين تباينات الخطأ ومستوى X، أن s_j^2 / X_j^2 مسـتقرة نسـبيا، وأن الأوزان $w_i = 1/X_i^2$ تبدو مناسبة.

أ - مستخدما الأوزان المقترحة، أوجد تقديرات المربعات الدنيـا المرجحـة لـِ β_0 و β_1 وانحرافاتها المعيارية المقدَّرة.

ب - ماذا تقدم المقارنة بين نتائجك في الجزء (أ) وتلك الـتي حصـل عليهـا الباحث؟ هل لاختيار الأوزان هنا تأثيرات مهمة؟ ناقش.

تمارين

(١١-٢٣) استنبط متوسط مربعات الخطأ في (11.40) .

(١١-٢٤) بالإشارة إلى التقديرات بطريقة الانحرافات المطلقة الدنيا لمثال شحوم الجسم على الصفحة (٥٢٧) ونعني $b_0 = -7.027$ ، $b_1 = 0.4173$ و $b_2 = 0.5203$.

أ - أوجد مجموع الانحرافات المطلقة عن القيم التوفيقية المستندة إلى تقديرات الانحرافات المطلقة الدنيا.

ب ـ ومن أجل تقديرات المربعات الدنيا لمعاملات الانحدار $b_0 = -19.174$، $b_1 = 0.2224$ و$b_2 = 0.6594$، أوجد مجموع الانحرافات المطلقة. هل هذا المجموع أكبر من المجموع الذي نحصل عليه في الجزء (أ)؟.

(١١-٢٥) (يحتاج إلى حساب التفاضل) استنبط المعادلات الناظمية للمربعات الدنيا المرجحة وذلك لتوفيق دالة الانحدار الخطية حيث $\sigma_i^2 = \sigma^2 X_i$ وσ^2 ثابت تناسب.

(١١-٢٦) عبّر عن مقدّرات المربعات الدنيا المرجحة b_1 في (11.56a) بدلالة الانحرافات $Y_i - \overline{X}_w$ و $X_i - \overline{X}_w$، حيث $\overline{Y}_w$ و $\overline{X}_w$ متوسطان مرجحان.

(١١-٢٧) بالعودة إلى **سرعة الآلة** مسألة (١١-٢٠)، أثبت عدديا أن تقديرات المربعات الدنيا المرجحة التي حصلت عليها في الجزء (د) مطابقة لتلك التي تحصل عليها مستخدما التحويل (11.67a) والمربعات الدنيا العادية.

(١١-٢٨) بالعودة إلى **التعلم بمساعدة الحاسب** مسألة (١١-٢١). أثبت عدديا أن تقديرات المربعات الدنيا المرجحة التي حصلت عليها في الجزء (د) مطابقة لتلك التي تحصل عليها مستخدما التحويل (11.67a) والمربعات الدنيا العادية.

(١١-٢٩) لنعتبر معيار المربعات الدنيا المرجحة (11.54) بالأوزان المعطاة في (11.64b) اكتب مصفوفة التباين ـ التغاير لحدود الخطأ عندما يكون $i = 1,...,4$ افترض $\sigma\{\varepsilon_i, \varepsilon_j\} = 0$ من أجل $i \neq j$.

(١١-٣٠) استنبط مصفوفة التباين ـ التغاير $\sigma^2\{\mathbf{b}\}$ في (11.62) الخاصة بمقدرات المربعات الدنيا المرجحة عندما تكون مصفوفة التباين ـ التغاير للمشاهدات Y_i هي $\sigma^2\mathbf{W}^{-1}$ حيث $\mathbf{W}$ معطاة في (11.59).

مشاريع

(١١-٣١) بالعودة إلى **ارتياح المريض** مسألة (٧-١٧).

أ ـ أوجد **معاملات انحدار** الحافة المعيارية المقدّرة، وعوامل تضخم التباين، وR^2 عندما يأخذ ثابت الانحياز c القيم التالية:

0.000، 0.005، 0.01، 0.02، 0.03، 0.04، و0.05.

ب ـ ارسم أثر الحافة من أجل قيم c المعطاة. هـل تُظهـر معـاملات انحـدار الحافة تغيرات مرموقة في جوار $c = 0$.

جـ ـ اقترح قيمة معقولة لثابت الانحياز c مبنية علـى أثـر الحافـة، وقيـم الـ VIF وR^2.

د ـ حوّل معاملات الانحدار المعيارية المقدَّرة التي اخترتهـا في الجـزء (جـ) عائدا إلى المتغيرات الأصلية واحسب القيم التوفيقية للمشـاهدات الـ 23، إلى أي حد تتشابه هذه القيم التوفيقيـة مـع تلـك الـتي حصلـت عليها بطريقة المربعات الدنيا العادية في المسألة (٧-١٧)ب؟

(١١-٣٢) بالعودة إلى **رواتب المختصين في الرياضيات** مسألة (٧ـ٢٠).

أ ـ أوجـد معـاملات انحـدار الحافـة المعياريـة المقـدَّرة، وعوامـل تضخـم التباين، وR^2 من أجل القيم التالية لثابت الانحياز:

$c = 0.000$ ،0.005 ،0.01 ،0.02 ،0.03 ،0.04 ،0.05

ب ـ ارسم أثر الحافة من أجل قيم c المعطاة. هل تظهـر معـاملات انحـدار الحافة تغيرات مرموقة في جوار $c = 0$ ؟

جـ ـ اقترح قيمة معقولة لثابت الانحياز مستندة إلى أثر الحافة، قيم الـ VIF وR^2.

د ـ حوّل معاملات الانحدار المعيارية المقدّرة التي اخترتهـا في الجـزء (جـ) عائدا إلى المتغيرات الأصلية، واحسب القيم التوفيقية للمشاهدات الـ 24، إلى أي حد تتشابه هـذه القيـم التوفيقيـة مـع تلـك الـتي حصلنـا عليها بطريقة المربعات الدنيا العادية في المسألة (٧ـ٢٠ب)؟.

(١١-٣٣) بالإشارة إلى مجموعة البيانات ***SENIC***.

أ ـ إحدر لوغاريتم طـول الإقامـة Y' علـى مخـاطرة العـدوى X_1، وعـدد الأسرّة X_2 ومتوسط التعداد اليومي X_3.

ب ـ أوجد الرواسب وحدد المشاهدات القاصية.

جـ ـ أوجد مصفوفة ارتباط المتغيرات X وعوامل تضخم التباين، ماذا تقترح هذه حول تأثيرات الخطية المتعددة؟

د ـ أوجد معاملات انحدار الحافة المقدّرة، وعوامل تضخم التباين وR^2 لقيم ثابت الانحياز c المعطاة في الجدول (١١-٦).

هـ ـ ارسم أثر الحافة وحدد قيمة معقولة لثابت الانحياز c مبنية على هذا الرسم، وعلى قيم الـ VIF وR^2.

(١١-٣٤) بالإشارة إلى مجموعة بيانات **SMSA**.

أ ـ إحدر عدد الأطباء العاملين Y على عدد أسِرّة المستشفى X_1، والدخل الشخصي الإجمالي X_2، والعدد الكلي للجرائم الخطرة X_3.

ب ـ أوجد الرواسب وحدد المشاهدات القاصية.

جـ ـ أوجد مصفوفة الارتباط للمتغيرات X وعوامل تضخم التباين، ماذا تقترح هذه حول تأثيرات الخطية المتعددة؟

د ـ أوجد معاملات انحدار الحافة المقدّرة، وعوامل تضخم التباين أو R^2 وذلك من أجل قيم ثابت الانحياز c المعطاة في الجدول (١١-٦).

هـ ـ ارسم أثر الحافة وحدد قيمة معقولة لثابت الانحياز c استنادا إلى هذا الرسم وإلى قيم الـ VIF وR^2.

(١١-٣٥) بالإشارة إلى **رواتب المختصين في الرياضيات**، مسألة (٧-٢٠).

أ ـ أوجد تقديرات الانحرافات المطلقة الدنيا للمعالم β_0، β_1، β_2 وβ_3.

ب ـ أوجد مجموع الانحرافات المطلقة عن القيم التوفيقية المستندة إلى تقديرات الانحرافات المطلقة.

جـ ـ أوجد مجموع الانحرافات المطلقة مستخدما دالة الانحدار المقدَّرة وفقا لطريقة المربعات الدنيا في المسألة (٧-٢٠)ب. هل هذا المجموع أكبر من المجموع الذي حصلت عليه في الجزء (ب)؟

(١١-٣٦) يُراد أخذ 5 مشاهدات مقابلة لـ $X = 10, 20, 30, 40, 50$ على الترتيب. ودالة الانحدار الصحيحة هي $E\{Y\} = 20 + 10X$ وحدود الخطأ مستقلة وتتوزع طبيعيا حيث $E\{\varepsilon_i\} = 0$ و $\sigma^2\{\varepsilon_i\} = 0.8X_i$

أ - ولّد مشاهدة عشوائية Y لكل مستوى من مستويات X. واحسب كلا من تقديرات المربعات الدنيا العادية والمرجحة لمعامل الانحدار β_1 في دالة الانحدار الخطية.

ب - أعد الجزء (أ) 200 مرة، مولدا أعدادا عشوائية جديدة في كل مرة.

جـ - احسب متوسط التباين للتقديرات المائتين لِـ β_1 وفق طريقة المربعات الدنيا العادية، وقم بالحسابات نفسها للتقديرات المائتين وفق طريقة المربعات الدنيا المرجحة.

د - هل يبدو مقدّار المربعات الدنيا العادية والمرجحة غير منحازين؟ اشرح. أي التقديرين يبدو أكثر دقة هنا؟ علّق.

بناء نموذج الانحدار

درسنا في الفصول السابقة كيفية توفيق نماذج انحدار بسيط ومتعدد، وكيفية القيام باستقراءات من هذه النماذج، وكيفية تشخيص شروط متنوعة تؤثر في صلاحية نموذج الانحدار التوفيقي.

ولأسباب تربوية، ناقشنا هذه المواضيع بمعزل عن بعضها البعض. ونحن في حاجة الآن لفحص كيفية تفاعلها فيما بينها في عملية بناء نموذج الانحدار. وفي هذا الفصل سنقدم أولا نظرة إجمالية لعملية بناء نموذج. ثم نناقش كل خطوة رئيسة من خطوات العملية بتفصيل أكبر. ومع القيام بذلك سنتعرض لإجراءات إضافية جديدة. وتنطوي إحدى المجموعات الجديدة من الإجراءات على تقنيات حاسوبية مفيدة لتحديد المتغيرات المستقلة التي سيشملها نموذج الانحدار. كما سنقدم أيضا عدة طرق للتحقق من صحة نموذج انحدار حال الانتهاء من تطويره وبنائه.

وعبر هذا الفصل سنستخدم المثال نفسه لتوضيح كل من خطوات عملية بناء النموذج، يتوج ذلك التحقق من صحة النموذج.

(١٢-١) نظرة إجمالية لعملية بناء نموذج

سنقدم في الشكل (١٢-١) استراتيجا لبناء نموذج انحدار، مخاطرين بما قد يبدو إفراطا في التبسيط. ويتضمن هذا الاستراتيج أربع خطوات:

١- جمع البيانات وإعدادها.

٢- تخفيض عدد المتغيرات المستقلة.

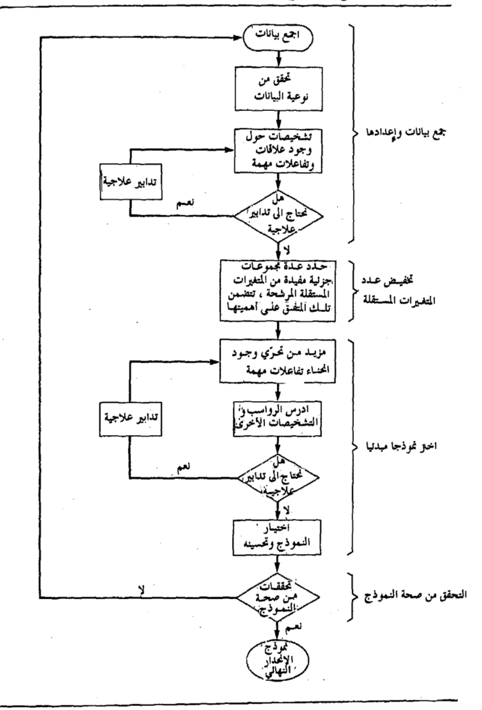
٣- تنقية النموذج واختيار المتغيرات المستقلة.

٤- التحقق من صحة النموذج.

جمع البيانات وإعدادها

في بعض الميادين يمكن أن تعين المعلومات النظرية في اختيار المتغيرات المستقلة التي ستُستخدم. وغالبا ما يمكن القيام بتجارب خاضعة لسيطرة المجرب، في مثل هذه الميادين، تزوده ببيانات يمكن، على أساسها، تقدير معالم الانحدار واختبار الشكل النظري لدالة الانحدار.

شكل (١٢-١) استراتيج لبناء نموذج انحدار



وفي العديد من الميادين الأخرى موضوع البحث، على أي حال، بما في ذلك العلوم الاجتماعية والسلوكية، الصحية والإدارية، يندر نسبيا وجود نماذج نظرية قابلة للاستخدام. ولمزيد من تعقيد الأمور، فقد تنطوي النماذج النظرية المتوافرة على متغيرات مستقلة غير قابلة للقياس مباشرة مثل الدخول المستقبلية للأسرة فوق السنوات العشر القادمة. وتحت مثل هذه الشروط يضطر الباحثون، في الغالب، إلى توقع متغيرات مستقلة يتصورون أنها يمكن أن تكون على صلة بالمتغير التابع المدروس. ومن الواضح أن مثل هذه المجموعة من المتغيرات المستقلة المفيدة يمكن أن تكون مجموعة كبيرة. وعلى سبيل المثال، فإن مبيعات شركة من غسالات الأطباق في منطقة يمكن أن تتأثر بحجم السكان، الدخل الفردي، النسبة المئوية للسكان الذين يعيشون في مناطق حضرية، النسبة المئوية من السكان تحت الخمسين سنة من العمر، النسبة المئوية من الأسر التي لها أطفال تحت سن المدرسة وإلخ، إلخ!

وبعد تجميع قائمة طويلة من المتغيرات المستقلة التي يُحتمل أن تكون مفيدة يمكن غربلة بعضها. إذ أولا قد لايكون متغير مستقل أساسيا في اعتبارات المسألة، وثانيا يمكن أن يكون المتغير المستقل خاضعا لأخطاء قياس كبيرة، و/ أو يمكن أن يكون ثالثا مجرّد تكرار أو صدى للمعلومات التي يقدمها متغير آخر في القائمة. ويمكننا إما حذف المتغيرات المستقلة التي لايمكن قياسها أو أن نستبدل بها متغيرات تقوم مقامها وتكون على ارتباط عال معها.

ويعتمد عدد المشاهدات التي نجمعها في دراسة انحدار على حجم الجملة من المتغيرات المستقلة التي قمنا بتجميعها في هذه المرحلة. فعندما يكون عدد المتغيرات المستقلة التي يمكن أن تكون مفيدة كبيرا نحتاج إلى مشاهدات أكثر مما لو كان عدد تلك المتغيرات صغيرا. وتعرض قاعدة باهم عامة أنه ينبغي توافر 6 إلى 10 مشاهدات، على الأقل، لكل متغير من مجموعة المتغيرات المقترحة. وتؤثر هذه الحاجة إلى مشاهدات أكثر من أجل مجموعات أكبر من المتغيرات المستقلة في استراتيجية بناء النموذج. وعلى سبيل المثال، قد يكون من الصعب تحرّي جميع التفاعلات بين متغيرين عندما يكون عدد المتغيرات المرشحة كبيرا، وربما كان ذلك ممكنا في حالة عدد صغير منها.

وحالما يتم جمع البيانات ينبغي القيام بالفحوصات والرسـومات اللازمـة لتحديـد الأخطاء الخـام في البيانـات بالإضافـة إلى القاصيـات. وعلـى وجـه الخصـوص، تكـون أخطاء البيانات منتشرة في مجموعـة بيانـات كبـيرة وينبغـي تصحيـح الأخطـاء أو حـل الصعوبات قبل البدء ببناء النموذج. وحيثما يكـون ممكنـا، ينبغي علـى البـاحث أن يراقـب بحذر عملية جمع البيانات ويديرها وذلك لتخفيض إمكانات وقوع أخطاء في البيانات.

وحالما نستكمل تهيئة البيانات التهيئة المناسبة يمكن أن تبدأ عملية النمذجة رسميا. وينبغي استخدام تشخيصات متنوعة لتحديد المتغيرات المستقلة المهمة، ولتحديد الصيـغ الدالّية التي ينبغي للمتغيرات أن تتخذها عند دخولها إلى النموذج، ولتحديد التفـاعلات المهمة. ورسوم الانتشار مفَيدة في تحديد العلاقات وقوة هذه العلاقات. ويمكــن توفيـق مختارات من المتغيرات المستقلة في دوال انحدار لاستكشـاف علاقـات وتفـاعلات قويـة ممكنة، وتحويلات. وهنا ينبغي لرسوم الراسب ورسوم الانحدار الجزئـي أن تلعـب دورا رئيسا، وبالطبع، وحيثما أمكن ذلك، ينبغي الاعتماد أيضا على المعرفة المسبقة للباحث وعلى خبرتـه العمليـة لاقـتراح تحويـلات مناسـبة وتفـاعلات يـرى تقصّيهـا. وجميـع إجراءات التشخيص المشروحة في الفصـول السـابقة ينبغـي لهـا أن تُستخدم كمصـادر معلومات في هذه المرحلة من بناء النموذج

تخفيض عدد المتغيرات المستقلة

وحالما يتخذ الباحث قراره الأولي حول الصيغ الدالية لعلاقـات الانحـدار (مـا إذا كان ينبغي أن تظهر متغيرات معينة في صيغة خطية، صيغة تربيعيـة، الخ) ومـا إذا كـان ينبغي للنموذج أن يشمل أية حدود تفاعل، فإن الخطوة التالية هي أن يختـار قليـلا مـن المجموعات الجزئية " الجيـدة " مـن المتغـيرات X. وينبغـي أن تتضمـن هـذه المجموعـات الجزئية ليس فقط المتغيرات المستقلة المرشحة في صيغ من المرتبة الأولى ولكنهـا تتضمـن أيضا أية حدود أخرى تربيعية أو منحنية نحتاجها وأية حدود تفاعلات ضرورية.

وسبب التركيز على مجموعات جزئية من جملـة المتغـيرات المستَقلة هـو أن عـدد المتغيرات المستقلة التي تبقى بعد الغربلة الابتدائية يبقى، في العادة، كبـيرا. وفضـلا عـن

ذلك، فكثيرا مايكون العديد من هذه المتغيرات المستقلة مرتبط بعضه ببعض ارتباطا عاليا، وبالتالي سيرغب الباحث عادة في تخفيض عدد المتغيرات المستقلة التي ستستخدم في النموذج النهائي. وهناك عدة أسباب لهذا، فمن الصعب صيانة نموذج انحدار بعدد كبير من المتغيرات المستقلة. وفضلا عن ذلك فإن العمل مع نماذج انحدار بعدد محدود من المتغيرات المستقلة سيكون أسهل وأيسر فهما. وأخيرا فإن وجود العديد من المتغيرات المستقلة المرتبطة فيما بينها ارتباطا عاليا يمكن أن يشكل إضافة بسيطة إلى قوة النموذج التنبؤية في الوقت الذي يزيد بشدة من تشتت المعاينة لمعاملات الانحدار، مما يقلص بدوره من القدرات الوصفية للنموذج، ويزيد من مشكلة الأخطاء الناتجة عن تدوير الأرقام العشرية (كما نوّهنا في الفصل الثامن).

واختيار مجموعات جزئية "جيدة" من المتغيرات المستقلة التي يُحتمل أن تكون مفيدة بغية احتوائها في نموذج الانحدار النهائي، وتحديد علاقات دالية وعلاقات تفاعل لهذه المتغيرات، يشكل عادة بعضا من مشاكل تحليل الانحدار الأكثر صعوبة. وبما أن لنماذج الانحدار استخدامات مختلفة فلا تشكل أية مجموعة جزئية واحدة من المتغيرات المستقلة، عادة، المجموعة الجزئية "الأفضل" على الدوام. وعلى سبيل المثال، بينما سيؤكد الاستخدام الوصفي لنموذج الانحدار، عادة، على دقة تقدير معاملات الانحدار، فسيركّز الاستخدام التنبؤي على أخطاء التنبؤ. وفي الغالب، فإن أفضل مايخدم هذه الأغراض المختلفة، هو مجموعات جزئية مختلفة من جملة المتغيرات المستقلة المرشحة. وحتى من أجل غرض معين، فقد وُجد، في الغالب، أن عدة مجموعات جزئية تتمتع "بالجودة" نفسها تقريبا، وذلك وفقا لمعيار معين، وينبغي القيام بالمفاضلة بين هذه المجموعات الجزئية "الجيدة" مستندين إلى اعتبارات إضافية.

ومن أجل بيانات المشاهدة ينبغي أن يتم اختيار قليل من المجموعات الجزئية المناسبة من المتغيرات المستقلة، للمفاضلة النهائية بينها، بعناية خاصة. ومع بيانات كهذه، فإن حذف متغيرات تفسيرية رئيسة يمكن أن يحطم بجدية القدرة التفسيرية للنموذج ويقود إلى تقديرات منحازة لمعاملات الانحدار، ولمتوسطات الاستجابة،

وللتنبؤات بمشاهدات جديدة، بالإضافة إلى تقديرات منحازة لتباين الخطأ. ويتصل الانحياز في هذه التقديرات بحقيقة أنه في بيانات مشاهده، يمكن أن تعكس حدود الخطأ في نموذج انحدار، يعاني نقصا في التوفيق، تأثيرات غير عشوائية للمتغيرات المستقلة التي لم تُستوعب في نموذج الانحدار. وأحيانا تدعى متغيرات مستقلة مهمة محذوفة، متغيرات تنبؤ مستترة.

وعلى الوجه الآخر، إذا تضمنت المجموعة الجزئية الكثير جدا من المتغيرات المستقلة، فسينتج نموذج كهذا يعاني من المبالغة في التوفيق، في الغالب، تباينات للمعالم المقدَّرة أكبر من تلك الناتجة عن نماذج أبسط.

والخطر الآخر عندما تكون البيانات بيانات مشاهدة هو أن المتغيرات المستقلة المهمة قد تُلحظ فوق أمداء ضيقة من القيم فقط. وكنتيجة لذلك فإن مثل هذه المتغيرات المستقلة المهمة قد تُحذف لمجرد وقوعها في العيّنة ضمن مدى ضيق من القيم، مما يجعلها تبدو غير مهمة إحصائيا.

والاعتبار الآخر في اختيار مجموعات جزئية من المتغيرات المستقلة، هو أنه ينبغي لهذه المجموعات الجزئية أن تكون صغيرة إلى الحد الذي يجعل تكاليف الصيانة معقولة، ويجعل تحليلها ميسّرا. ومع ذلك ينبغي لها من جهة أخرى أن تكون كبيرة إلى الحد الذي يجعل من الممكن القيام بوصف مناسب وتحكم وتنبؤ مناسبين.

وقد طُوِّرت أساليب حاسوبية متنوعة لمساعدة الباحث في تخفيض عدد المتغيرات المستقلة التي سيعتبرها في نموذج انحدار، وذلك عندما تكون هذه المتغيرات مرتبطة فيما بينها. وسنقدم أسلوبين من هذه الأساليب في هذا الفصل. والأسلوب الأول، وهو أسلوب عملي من أجل جملة من المتغيرات المستقلة صغيرة في حجمها أو معتدلة الحجم، يدرس جميع نماذج الانحدار الممكنة التي نستطيع تطويرها من جملة المتغيرات المستقلة المرشحة، ويحدد مجموعات جزئية من المتغيرات المستقلة هي المجموعات "الأجود" وفقا لمعيار يحدده الباحث. ويستخدم الأسلوب الثاني طرق بحث آلية للوصول إلى مجموعة جزئية واحدة من المتغيرات المستقلة. ويُوصى بهذا الأسلوب مبدئيا في حالة عمليات تخفيض تنطوي على مجموعات كبيرة من المتغيرات المستقلة.

وفي الوقت الذي يمكن أن تقدم فيه الأساليب الحاسوبية مساعدة كبيرة، من حيث تحديد مجموعات جزئية مناسبة لدراستها دراسة مفصلة ونهائية، إلا أنه لابد لعملية تطوير نموذج انحدار مفيد أن تكون ذرائعية وأن تحتاج إلى الاستعانة بجرعات كبيرة من الحكم الشخصي. وينبغي للمتغيرات المستقلة التي تُعتبر أساسية أن تأخذ مكانها في نموذج الانحدار قبل طلب أي مساعدة حاسوبية. وفضلا عن ذلك فإنه لابد من تتمات للأساليب الحاسوبية التي تقدم مجموعة جزئية واحدة فقط من المتغيرات المستقلة "كأفضل" مجموعة جزئية، بحيث نتمكن أيضا من أخذ مجموعات جزئية أخرى في الاعتبار قبل تقرير الشكل النهائي لنموذج الانحدار.

تعليقات

١- تلغي التجارب المصممة بعناية، عادة، العديد من المشاكل المتصلة باختيار مجموعات جزئية "جيدة" من المتغيرات المستقلة. وعلى سبيل المثال، يمكن جعل تأثيرات متغيرات التنبؤ المستترة أصغر مايمكن باستخدام العشوائية. وبالإضافة إلى ذلك، يمكن اختيار أمداء مناسبة لمتغيرات التنبؤ، كما يمكن إلغاء الارتباطات بين متغيرات التنبؤ من خلال اختيارات مناسبة لمستوياتها.

٢- في كثير من الأحيان سيغربل باحث غير واع مجموعة من المتغيرات المستقلة بتوفيق نموذج انحدار يتضمن المجموعة بكاملها من المتغيرات X المرشحة ثم يلغي ببساطة تلك المتغيرات التي تكون القيمة المطلقة للإحصاءة t^* في (8.23) من أجلها صغيرة حيث:

$$t_k^* = \frac{b_k}{s\{b_k\}}$$

وكما نعلم من الفصل الثامن، فقد تقود هذه الطريقة إلى إلغاء متغيرات مستقلة مهمة ولكن مرتبطة فيما بينها. ومن الواضح أن أسلوب البحث الجيد يجب أن يكون قادرا على تناول متغيرات مستقلة مهمة ومرتبطة فيما بينها بطريقة لاتلغيها جميعا من النموذج.

اختيار وتحسين النموذج

بعد تخفيض عدد المتغيرات المستقلة بصورة ناجحة يصل الباحث عادة إلى عدد صغير من نماذج الانحدار التي يُحتمل أن تكون "جيدة" وكل منها يتضمن تلك المتغيرات المستقلة المعروفة بأنها متغيرات أساسية. ومن المستحسن في هذه المرحلة القيام بتحقيقات أكثر تفصيلا عن تأثيرات التفاعل والانحناء. وتكون رسوم الراسب ورسوم الانحدار الجزئي معيْنة في تقرير ما إذا كان يمكن تفضيل نموذج على آخر، وبالإضافة إلى ذلك تكون الفحوصات التشخيصية الموصوفة في الفصل الحادي عشر مفيدة لتحديد المشاهدات القاصية المؤثرة، والخطية المتعددة، إلخ.

وأخيرا وبعد فحص شامل وتدابير علاجية مختلفة، كالتحويلات مثلا، يقرر الباحث اختيار أحد نماذج الانحدار كأفضل نموذج. وأحد الممارسات الإحصائية المحمودة عند هذه النقطة هي التحقق من صحة النموذج.

التحقق من صحة نموذج

تشير صحة النموذج إلى استقرار ومعقولية معاملات الانحدار، وإلى قابلية دالة الانحدار للاستخدام ودرجة نجاحها وإلى إمكانية تعميم الاستقراءات المستخلصة من تحليل الانحدار. والتحقق من صحة نموذج هي جزء ضروري ومفيد في عملية بنائه. وسنصف لاحقا، في هذا الفصل، عدة طرق لتثمين صحة نموذج.

(١٢-٢) إعداد البيانات

ولتوضيح إجراءات بناء نموذج التي ناقشناها في هذا الفصل، سنستخدم مثالا بسيطا نسبيا فيه أربعة متغيرات مستقلة مرشحة. ومع تحديد عدد صغير من المتغيرات المستقلة المرشحة سنكون قادرين على شرح الإجراءات دون أن نغمر القارىء بفيض من مُخرجات الحاسب.

وقد ذكرنا سابقا بعضا من خطوات إعداد البيانات التي يأتي دورها في بداية عملية بناء النموذج. ونوضح هذه الخطوات، الآن بدلالة مثال الوحدة الجراحية.

مثال

تهتم وحدة جراحة في مستشفى بالتنبؤ بنسبة الشفاء لمرضى يخضعون لنوع معين من جراحة الكبد. وقد توافر للتحليل 54 مريضا اختيروا عشوائيا. ومن سجل كل مريض استُخلصت المعلومات التالية من التقويم الذي يسبق العملية:

X_1 درجة تخثر الدم.

X_2 دليل الإنذار القياسي، بما في ذلك عمر المريض.

X_3 درجة اختبار وظيفة الأنزيم

X_4 درجة اختبار وظيفة الكبد

وهي تشكل المتغيرات المستقلة المرشحة لنموذج انحدار تنبؤي. والمتغير التابع هو الفترة التي يعيشها المريض بعد الجراحة، والتي تجري معرفتها من متابعة أحوال المريض. والبيانات عن المتغيرات المستقلة المرشحة والمتغير التابع مقدمة في الجدول (١٢-١). وقد غُربلت هذه البيانات وروجعت بصورة مناسبة من أجل الأخطاء.

وبما أن جملة المتغيرات المستقلة صغيرة، فمن الممكن في هذه المرحلة من إعداد البيانات القيام باستكشاف كامل إلى حد ما، لتأثيرات علاقات وتفاعلات قوية محتملة. وقد تمّ إعداد رسوم جذع وورقة لكل من المتغيرات المستقلة (غير مبينة هنا). وقد أوضحت هذه الرسوم بجلاء أن الحالة 28 هي مشاهدة قاصية في X_1 وأن الحالة 38 مشاهدة قاصية في X_2، وأن الحالتين 17 و32 مشاهدتان قاصيتان في X_3، وأن الحالات 28، 31 و43 مشاهدات قاصية في X_4 وهكذا تأهب الباحث لفحص نفوذ هذه الحالات لاحقا.

وقد تم توفيق نموذج انحدار من المرتبة الأولى يرتكز إلى المتغيرات المستقلة جميعها. ورسم الاحتمال الطبيعي لرواسب هذا النموذج التوفيقي مبيّن في الشكل (١٢-٢)أ. وهو يقترح حيدانا عن الطبيعية. ومعامل الارتباط بين الرواسب المرتّبة وقيمها المتوقعة تحت الطبيعية هو 0.826 فقط ؛ وبالعودة إلى الجدول (٤-٣) نجد أن هذا الرقم يدعم الاستنتاج بأن حدود الخطأ لاتتبع التوزيع الطبيعي.

جدول (١٢-١) المتغيرات المستقلة المرشحة والمتغير التابع ـ مثال وحدة الجراحة.

رقم الحالة i	درجة تخثر الدم X_{i1}	درجة اعتبار وظيفة الأنزيم X_{i2}	اختبار وظيفة الانزيم X_{i3}	اختبار وظيفة الكبد X_{i4}	فترة الحياة بعد الجراحة Y_i	$Y_i' = \log_{10} Y_i$
1	6.7	62	81	2.59	200	2.3010
2	5.1	59	66	1.70	101	2.0043
3	7.4	57	83	2.16	204	2.3096
4	6.5	73	41	2.01	101	2.0043
5	7.8	65	115	4.30	509	2.7067
6	5.8	38	72	1.42	80	1.9031
7	5.7	46	63	1.91	80	1.9031
8	3.7	68	81	2.57	127	2.1038
9	6.0	67	93	2.50	202	2.3054
10	3.7	76	94	2.40	203	2.3075
11	6.3	84	83	4.13	329	2.5172
12	6.7	51	43	1.86	65	1.8129
13	5.8	96	114	3.95	830	2.9191
14	5.8	83	88	3.95	330	2.5185
15	7.7	62	67	3.40	168	2.2253
16	7.4	74	68	2.40	217	2.3365
17	6.0	85	28	2.98	87	1.9395
18	3.7	51	41	1.55	34	1.5315
19	7.3	68	74	3.56	215	2.3324
20	5.6	57	87	3.02	172	2.2355
21	5.2	52	76	2.85	109	2.0374
22	3.4	83	53	1.12	136	2.1335
23	6.7	26	68	2.10	70	1.8451
24	5.8	67	86	3.40	220	2.3424
25	6.3	59	100	2.95	276	2.4409
26	5.8	61	73	3.50	144	2.1584
27	5.2	52	86	2.45	181	2.2577
28	11.2	76	90	5.59	574	2.7589
29	5.2	54	56	2.71	72	1.8573
30	5.8	76	59	2.58	178	2.2504
31	3.2	64	65	.74	71	1.8513
32	8.7	45	23	2.52	58	1.7634
33	5.0	59	73	3.50	116	2.0645
34	5.8	72	93	3.30	295	2.4698
35	5.4	58	70	2.64	115	2.0607
36	5.3	51	99	2.60	184	2.2648
37	2.6	74	86	2.05	118	2.0719

جدول ١٢–١ (تتمة)

رقم الحالة i	درجة تخثر الدم X_{i1}	درجة اعتبار وظيفة الأنزيم X_{i2}	اختبار وظيفة الانزيم X_{i3}	اختبار وظيفة الكبد X_{i4}	فترة الحياة بعد الجراحة Y_i	$Y_i' = \log_{10} Y_i$
38	4.3	8	119	2.85	120	2.0792
39	4.8	61	76	2.45	151	2.1790
40	5.4	52	88	1.81	148	2.1703
41	5.2	49	72	1.84	95	1.9777
42	3.6	28	99	1.30	75	1.8751
43	8.8	86	88	6.40	483	2.6840
44	6.5	56	77	2.85	153	2.1847
45	3.4	77	93	1.48	191	2.2810
46	6.5	40	84	3.00	123	2.0899
47	4.5	73	106	3.05	311	2.4928
48	4.8	86	101	4.10	398	2.5999
49	5.1	67	77	2.86	158	2.1987
50	3.9	82	103	4.55	310	2.4914
51	6.6	77	46	1.95	124	2.0934
52	6.4	85	40	1.21	125	2.0969
53	6.4	59	85	2.33	198	2.2967
54	8.8	78	72	3.20	313	2.4955

وقد تم توفيق نماذج أخرى متنوعة من المرتبة الأولى، وإعداد العديد من رسوم الراسب. وأحد الرسوم المفيدة بوجه خاص كان رسم الرواسب، الناتج عن توفيق Y على X_2 و X_3، في مقابل حد التفاعل X_2X_3 وهذا الرسم مبيّن في الشكل (١٢–٢)ب. لاحظ تأثيرات التفاعل القوية التي يقترحها هذا الرسم.

ولجعل توزيع حدود الخطأ أقرب إلى الطبيعي، ولرؤية ما إذا كان التحويل نفسه يمكن أن يخفض تأثير التفاعل X_2X_3، قام الباحث بالتحويل اللوغاريتمي $Y' = log_{10}Y$. والبيانات الخاصة بالمتغير التابع بعد تحويله معطاة في الجدول (١٢–١). ويبين الشكل (١٢-٢)جـ رسم الاحتمال الطبيعي للرواسب عند حدر Y' على جميع المتغيرات المستقلة في نموذج من المرتبة الأولى، ومعامل الارتباط بين الرواسب المرتّبة وقيمها المتوقعة تحت الطبيعية هو 0.959. ومن الواضح أن التحويل قد أعان في جعل توزيع حدود الخطأ أكثر طبيعية. وفضلا عن ذلك، فقد كان التحويل مُعينا أيضا لتخفيض أثر حد التفاعل

بين X_2 وX_3. ويبين الشكل (١٢-٢)د رسم الرواسب، عند حدر Y' على X_2 و X_3 في مقابل X_2X_3. والمؤشر الذي يقدمه رسم الراسب هذا عن قوة تأثير التفاعل هـو مؤشـر أقل بكثير. ولم يقدم أي مـن رسـوم الراسـب في مقـابل حـدود التفـاعل مؤشـرا علـى وجود أي تأثير تفاعل قوي.

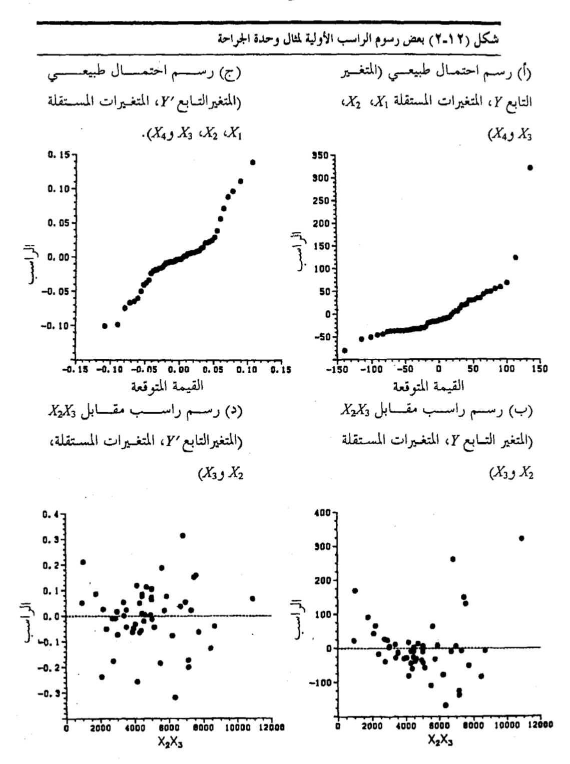
وقد حصل الباحث على مصفوفة الارتباط متضمنة المتغير Y بعد التحويل ؛ وهي معطاة في الجدول (١٢-٢). مع حذف الحـدود المكـررة. وبالإضافـة إلى ذلـك فقـد تم الحصول على رسوم انتشار لِـ Y' في مقابل كل متغير مستقل ولكل زوج من المتغيرات المستقلة. والشكل (١٢-٣) هو مثال توضيحي يبين رسم انتشار Y' في مقابل X_1.

والجـدول (١٢ــ٢) بالإضافـة إلى رسـوم الانتشـار المختلفـة، ورسـوم الراسـب، ورسوم الانحدار الجزئي (غير مبينـة هنـا، باسـتثناء رسـم الانتشـار في الشـكل ١٢ــ٣) تشيرجميعها إلى صلة خطية بين كل مـن المتغـيرات المستقلة وبـين Y'، ودرجـة الصلـة الخطية هي الأشد في حالة X_4 والأضعف في حالة X_1، وتبـين مصفوفـة الارتبـاط أيضـا معاملات الارتباط فيما بين المتغيرات المستقلة المرشحة. وعلى وجـه الخصـوص، يُظهـر X_4 ارتباطا مرتفعا باعتدال مع كل من X_1، X_2 وX_3.

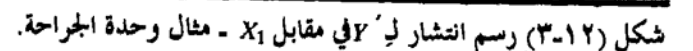
وعلى أساس من هذه التحليلات، قرر الباحث استخدام كمتغير تـابع $Y' = log_{10}Y$، في هذه المرحلة من بناء النموذج، وتمثيل المتغيرات المستقلة وفق حدود خطية، مستبعدا أي حدود تفاعل.

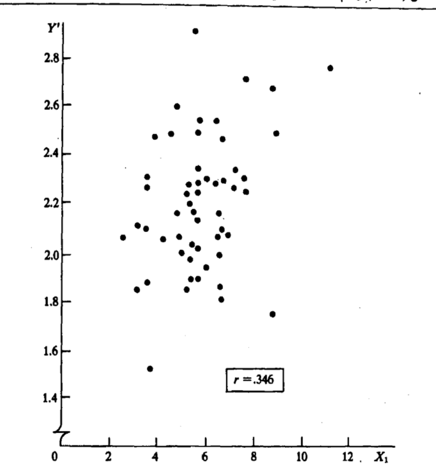
جدول (١٢-٢) مصفوفة الارتباط لمثال وحدة الجراحة

X_4	X_3	X_2	X_1	Y'	
.726	.665	.593	.346	1.000	Y'
.502	-.150	.090	1.000		X_1
.369	-.024	1.000			X_2
.416	1.000				X_3
1.000					X_4



شكل (١٢-٢) بعض رسوم الراسب الأولية لمثال وحدة الجراحة


شكل (١٢-٣) رسم انتشار لِـ Y' في مقابل X_1 - مثال وحدة الجراحة.



(١٢-٣) طريقة جميع الانحدارات الممكنة لتخفيض المتغيرات

تستدعي طريقة جميع الانحدارات الممكنة للاختيار اعتبار جميع نمـاذج الانحـدار الممكنـة المتضمنة للمتغيرات X المرشحة وتحديد عدد قليل من المجموعات الجزئية "الجيدة" وفقـا لمعيار ما. وعند استخدام هذه الطريقة في الاختيـار في مثـال الوحـدة الجراحيـة، مثـلا، فسنعتبر16 من نماذج الانحدار المختلفة كما هــو مبيّـن في الجـدول (١٢-٣). ويوجـد، أولا، نموذج الانحدار الذي لايحوي متغيرات X، أي النموذج $Y_i = \beta_0 + \varepsilon_i$. وهناك بعد ذلك، نماذج الانحدار بمتغير X واحد، (X_1, X_2, X_3, X_4)، ونماذج الانحـدار بمتغـيري X $(X_1$ وX_2، X_1 و X_3، X_1 و X_4، X_2 وX_3، X_2 وX_4، X_3 و$X_4)$ وهلم جرا.

والغرض من أسلوب جميع الانحدارات الممكنة هو تحديد مجموعة صغيرة من نماذج الانحدار "الجيدة" وفقا لمعيار محدد بحيث يمكن القيام بفحص تفصيلي لهذه النماذج، مما يقود بدوره إلى اختيار نموذج الانحدار النهائي الذي سيجري استخدامه. وتحت معظم الظروف، قد يكون من المستحيل على محلل القيام بفحص تفصيلي لجميع نماذج الانحدار الممكنة. وعلى سبيل المثال، عندما يوجد 10 متغيرات مستقلة مرشحة فسيكون هناك $2^{10} = 1024$ من نماذج الانحدار الممكنة. (يستند هذا الحساب على حقيقة أن كل متغير مستقل مرشح إما أن يكون ضمن النموذج أو خارجه)، ومع توافر الحاسبات ذات السرع العالية اليوم، فإن تشغيل جميع نماذج الانحدار الممكنة لعشر متغيرات X مرشحة لايستهلك الكثير من الوقت. ومع ذلك فإن فحص مجلد يتضمن 1024 نموذجا مرشحا يمكن أن يشكل عملا مضنيا بالنسبة لمحلل بيانات.

وبالتالي فإنه ما لم تكن جملة المتغيرات X المرشحة صغيرة جدا فسيركز الباحث على قليل فقط من مجموعة كل نماذج الانحدار الممكنة. ويمكن أن يتضمن هذا العدد المحدود من النماذج 5 أو 10 مجموعات جزئية "جيدة" وفقا لمعيار محدد، مما يسمح للباحث عندئذ أن يدرس بعناية نماذج الانحدار هذه ليختار من بينها النموذج النهائي الذي سيستخدمه.

وفي طريقة جميع الانحدارات الممنكة للاختيار يمكن استخدام معايير مختلفة لمقارنة نماذج الانحدار. وسنناقش أربعا منها R_p^2 ، MSE_p، Cp و$PRESS_P$. وقبل القيام بذلك، نحتاج إلى اعتماد بعض الرموز، فلنرمز لعدد المتغيرات X المرشحة في الجملة بـ P-1. ونفترض، عبر هذا الفصل، أن جميع نماذج الانحدار تتضمن حد الجزء المقطوع β_0. وبالتالي تتضمن دالة الانحدار التي تحوي جميع المتغيرات X المرشحة P معلمة، وتتضمن الدالة التي تخلو من المتغيرات X معلمة واحدة β_0.

وسنرمز لعدد المتغيرات X في مجموعة جزئية، على الدوام، بـ p -1، وبحيث توجد p معلمة في دالة الانحدار الخاصة بهذه المجموعة الجزئية من المتغيرات X وهذا يكون:

$$1 \le p \le P \qquad (12.1)$$

ويفترض أسلوب جميع الانحدارات الممكنة أن عـدد المشـاهدات n يتجـاوز أكـبر عدد من المعالم المرشحة:

$$n > P \tag{12.2}$$

وكما ذكرنا سابقا، فمن المستحسن جدا، في الحقيقة، أن يكون n أكبر بكثير من P، بحيث يمكن الحصول على نتائج سليمة.

معيار الـ R_p^2

ويستدعي معيار الـ R_p^2 فحص معامل التحديـد المتعـدد R^2 المعـرف في (7.35)، كي نختار عدة مجموعات جزئية "جيدة" من المتغيرات X. ونضع عدد المعالم في نمـوذج الانحدار كدليل لِ R^2. وهكذا يشير R_p^2 إلى وجود p معلمة، أو $p - 1$ متغير تنبؤ، في دالة الانحدار التي استند اليها R_p^2.

وبما أن R_p^2 هي نسبة لمجموعي مربعات:

$$R_p^2 = \frac{SSR_p}{SSTO} = 1 - \frac{SSE_p}{SSTO} \tag{12.3}$$

والمقام ثابت من أجل نماذج الانحدار الممكنة جميعها، فإن R_p^2 تتغير عكسيا مع مجموع مربعـات الخطـأ SSE_p. ولكننـا نعلـم أن SSE_p لا يمكـن لـه أن يـزداد عندمـا يتضمـن النموذج مزيدا من المتغيرات X الإضافيـة، وهكـذا سـيكون R_p^2 أكـبر مـايمكن عندمـا يتضمن نموذج الانحدار جميع المتغيرات X المرشحة وعددها $P - 1$ وبالتـالي لايمكـن أن يكون سبب استخدام المعيار R_p^2 في أسلوب جميع الانحدارات الممكنة هـو جعـل R_p^2 أعظم ما يمكن. وإنما القصد هو إيجاد الوضع الذي تصبح إضافة المزيد من المتغـيرات X عنده غير ذات شأن باعتبارها تؤدي إلى زيادة صغيرة جدا في R_p^2. وفي الغالب نصـل إلى مثل هذا الوضع عندما يتضمن نموذج الانحدار عددا محدودا فقط مـن المتغـيرات X. ومن الواضح، أن تحديد الوضع الذي يبدأ عنده العـائد بالتلاشـي هـي مسـألة اجتهـاد شخصي.

مثال. يبين الجدول (١٢-٣) لمثال الوحدة الجراحية، في الأعمدة (١)، (٢)و (٣) علـى الترتيب، عدد المعالم في دالة الانحدار، درجات الحرية الموافقة لمجمـوع مربعـات الخطـأ،

ومجموع مربعات الخطأ، وذلك لكل نموذج انحدار ممكن. وفي العمود (٤) نجد قيم R_p^2، وقد تمّ الحصول على النتائج من سلسلة من التشغيلات للحاسب.

جدول (١٢-٣). قيم R_p^2، MSE_p، C_p و $PRESS_p$ لجميع نماذج الانحدار المحتملة - مثال الوحدة الجراحية.

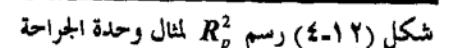
(٧) $PRESS_p$	(٦) Cp	(٥) $MSEp$	(٤) R_p^2	(٣) $SSEp$	(٢) df	(١) p	المتغيرات في النموذج X
4.1241	1,721.6	.0750	0	3.9728	53	1	
3.8084	1,510.8	.0672	.120	3.4961	52	2	X_1
2.8627	1,100.1	.0495	.352	2.5763	52	2	X_2
2.4268	939.0	.0426	.442	2.2153	52	2	X_3
2.0292	788.2	.0361	.527	1.8776	52	2	X_4
2.6388	948.7	.0438	.438	2.2325	51	3	X_1, X_2
1.6095	580.2	.0276	.646	1.4072	51	3	X_1, X_3
2.1203	789.4	.0368	.528	1.8758	51	3	X_1, X_4
.8352	283.7	.0146	.813	.7430	51	3	X_2, X_3
1.5833	573.5	.0273	.650	1.3922	51	3	X_2, X_4
1.4287	507.9	.0244	.687	1.2453	51	3	X_3, X_4
.1405	3.1	.00220	.972	.1099	50	4	X_1, X_2, X_3
1.6513	574.8	.0278	.650	1.3905	50	4	X_1, X_2, X_4
1.3286	452.0	.0223	.719	1.1156	50	4	X_1, X_3, X_4
.5487	161.7	.00930	.883	.4652	50	4	X_2, X_3, X_4
.1456	5.0	.00224	.972	.1098	49	5	X_1, X_2, X_3, X_4

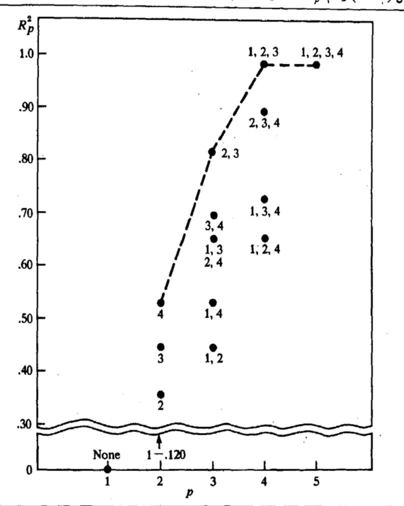
وعلى سبيل المثال، عندما يكون X_4 هو المتغير X الوحيد في نموذج الانحدار، نجد:

$$R_2^2 = 1 - \frac{SSE(X_4)}{SSTO} = 1 - \frac{1.8776}{3.9728} = .527$$

ونلاحظ أن $SSTO = SSE_1 = 3.9728$.

وقيم R_p^2 مرسومة في الشكل (١٢-٤). والقيمة العظمى لـ R_p^2 من أجل المجموعات الجزئية من $p - 1$ متغير تنبؤ، ونرمز لها بـ $max(R_p^2)$، تبدو في قمة البيان لكل قيمة من قيم p. وهذه النقاط موصولة بخطوط متقطعة لتبيان أثر إضافة مزيد من المتغيرات X. ويوضح الشكل (١٢-٤) أنه بعد أن يصبح النموذج متضمنا لثلاثة متغيرات X فإن الزيادة الحاصلة في $max\ (R_p^2)$ هي زيادة طفيفة. وبالتالي يبدو استخدام المجموعة الجزئية (X_1, X_2, X_3) في نموذج الانحدار أمرا منطقيا وفقا لمعيار الـ R_p^2.


شكل (١٢-٤) رسم R_p^2 لمثال وحدة الجراحة



لاحظ أن المتغير X_4 الذي ينفرد بأعلى ارتباط مع المتغير التابع غير موجود في النمـاذج التي يقترحها $max(R_p^2)$ سواء من أجـل $p = 3$ و$p = 4$، ممـا يشـير إلى أن X_1، X_2 و X_3 تحتوي كثيرا من المعلومات التي يقدمهـا X_4، وإذا رغبنـا الاحتفـاظ بـ X_4 في النمـوذج، مـع اقتصار المجموعة الجزئية التي يتضمنها النموذج على ثلاثة متغـيرات X فينبغـي عندئـذ اعتبـار المجموعة الجزئية (X_2, X_3, X_4) المجموعة التالية في الأفضلية وفقا لمعيار الـ R_p^2 من أجل $p = 4$. ومعامل الانحدار المتعدد المقابل لهذه المجموعـة الجزئيـة، وهـو $R_4^2 = 0.883$ ، سيكون أصغـر بقليل من $R_4^2 = 0.972$ المقابل للمجموعة الجزئية (X_1, X_2, X_3).

معيار MSE_p أو R_a^2

بما أن R_p^2 لايأخذ في الاعتبار عدد المعالم في نموذج الانحدار، وبما أن $max(R_p^2)$ لايمكن أن تتناقص أبدا مع زيادة p، فقد اقتُرح استخدام معامل التحديد المتعدد المعدّل R_a^2 المذكور في (7.37):

$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO} = 1 - \frac{MSE}{\frac{SSTO}{n-1}} \qquad (12.4)$$

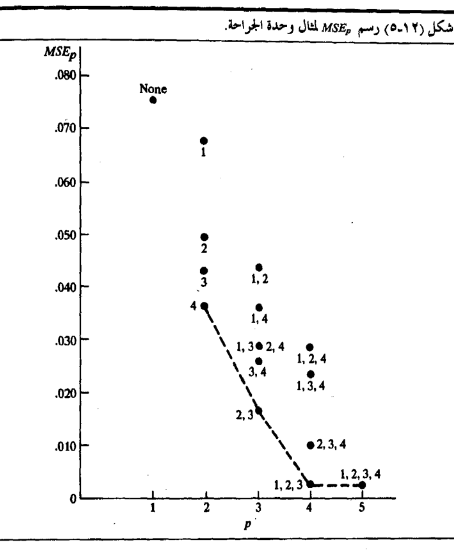
كمعيار يأخذ عدد معالم النموذج في الاعتبار وذلك من خلال درجات الحرية. ومن (12.4) يمكن رؤية أن R_a^2 يزداد إذا وفقط إذا تناقص MSE، ذلك لأن $SSTO\ /(n\text{-}1)$ مثبت من أجل مشاهدات Y معطاة. وبالتالي فإن R_a^2 وMSE هما معياران متكافئان. وسنعتبر هنا المعيار MSE_p إذ يمكن لِـ $min(MSE_p)$ في الحقيقة، أن يزداد مع زيادة p، وذلك عندما يصبح التخفيض في SSE_p أصغر من أن يعوّض خسارة درجة حرية إضافية. ويسعى مستخدمو المعيار MSE_p إلى إيجاد عدد بسيط من المجموعات الجزئية التي يكون MSE_p، من أجلها، في قيمته الدنيا، أو قريبا جدا من القيمة الدنيا، وبحيث لايصبح لإضافة المزيد من المتغيرات شأن يذكر.

مثال. قيم الـ MSE_p لجميع نماذج الانحدار الممكنة في مثال وحدة الجراحة معروضة في العمود الخامس من الجدول (١٢-٣). وعلى سبيل المثال، إذا اقتصر نموذج الانحدار على X_4، فلدينا:

$$MSE_2 = \frac{SSE(X_4)}{n-2} = \frac{1.8776}{52} = 0.0361$$

ويتضمن الشكل (١٢-٥) رسم الـ MSE_p لمثال وحدة الجراحة. وقد وصلنا القيم $min(MSE_p)$ من أجل كل p بخطوط متقطعة. والقصة التي يرويها الشكل (١٢-٥) مماثلة تماما لتلك التي يرويها الشكل (١٢-٤). إذ تبدو المجموعة الجزئية (X_1, X_2, X_3) معقولة.

شكل (١٢-٥) رسم MSE_p لمثال وحدة الجراحة.



وفي الحقيقة، فإن متوسط مربعات الخطأ الذي تحرزه هذه المجموعة الجزئية هو عمليا نفس ماتحرزه المجموعة الجزئية (X_1, X_2, X_3, X_4) التي تستخدم جميع المتغيرات X المرشحة.

وإذا رغبنا أن يتضمن النموذج X_4 مع $p = 4$، فينبغي أخذ المجموعة الجزئية (X_2, X_3, X_4) في الاعتبار التي يقابلها $MSE_4 = 0.009$ وهو إلى حد ما أعلى من $MSE_4 = 0.002$ المقابل للمجموعة الجزئية (X_1, X_2, X_3).

المعيار C_p

يهتم هذا المعيار بمتوسط مربعات الخطأ الكلي للقيم التوفيقية الـ n وذلك لكل نموذج انحدار قائم على مجموعة جزئية، وتنطوي فكرة متوسط مربعات الخطأ على مركبة انحياز ومركبة خطأ عشوائي. وقد عرفنا في (11.40) متوسط مربعات الخطأ لمعامل انحدار مقدَّر، ويتعلق متوسط مربعات الخطأ هنا بالقيم التوفيقية $\hat{Y}_i$ لنموذج الانحدار المستخدم. ومركبة الانحياز للقيمة التوفيقية الـ i، ونعني $\hat{Y}_i$ هي:

$$E\{\hat{Y}_i\} - \mu_i \qquad (12.5)$$

حيث $E\{\hat{Y}_i\}$ هو توقع القيمة التوفيقية الـ i لنموذج الانحدار المدروس و μ_i هو متوسط الاستجابة الصحيح. والمركبة العشوائية لِـ $\hat{Y}_i$ هي، ببساطة تباينها $\sigma^2\{\hat{Y}_i\}$. وعندئذ يكون متوسط مربعات الخطأ لِـ $\hat{Y}_i$ هو مربع الانحياز مضافا إليه التباين:

$$\left(E\{\hat{Y}_i\} - \mu_i\right)^2 + \sigma^2\{\hat{Y}_i\} \qquad (12.6)$$

والمجموع الكلي لمتوسط مربعات الخطأ لجميع القيم التوفيقية $\hat{Y}_i$ وعددها n ، هو مجموع n من متوسطات مربعات الخطأ المفردة كتلك المعطاة في (12.6):

$$\sum_{i=1}^{n}\left[\left(E\{\hat{Y}_i\} - \mu_i\right)^2 + \sigma^2\{\hat{Y}_i\}\right] = \sum_{i=1}^{n}\left(E\{\hat{Y}_i\} - \mu_i\right)^2 + \sum_{i=1}^{n}\sigma^2\{\hat{Y}_i\} \qquad (12.7)$$

وقيمة المعيار، ونرمز لها بـ Γ_p، هي ببساطة متوسط مربعات الخطأ الكلي في (12.7) مقسوما على التباين الفعلي للخطأ σ^2:

$$\Gamma_p = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}\left(E\{\hat{Y}_i\} - \mu_i\right)^2 + \sum_{i=1}^{n}\sigma^2\{\hat{Y}_i\}\right] \qquad (12.8)$$

ومن المفترض أن النموذج المتضمن لجميع المتغيرات X المرشحة، وعددها $P - 1$، قد اختيرت بعناية بحيث يكون $MSE(X_1,..., X_{P-1})$ مقدرا غير منحاز لِـ σ^2. ويمكن عندئذ تبيان أن مقدِّر Γ_p هو C_p حيث:

$$C_p = \frac{SSE_p}{MSE(X_1,\ldots,X_{P-1})} - (n-2p) \qquad (12.9)$$

وحيث MSE_p مجموع مربعات الخطأ لنموذج الانحدار الجزئي الذي تم توفيقه بـ p من المعالم (أي $p - 1$ متغير تنبؤ).

وعندما لايوجد في نموذج انحدار يتضمن $p-1$ متغير تنبؤ بحيث يكون $E\{\hat{Y}_i\} \equiv \mu_i$ فإن القيمة المتوقعة لِـ C_p تساوي تقريبا p:

$$E\{C_p \mid E\{\hat{Y}_i\} \equiv \mu_i\} \cong p \tag{12.10}$$

وهكذا إذا رسمنا قيم C_p من أجل جميع نماذج الانحدار الممكنة في مقابل p، فستنحو تلك النماذج ذات الانحياز الطفيف إلى الوقوع قرب الخط $C_p = p$ وستنحو النماذج ذات الانحياز الكبير إلى الوقوع فوق هذا الخط وبعيدا عنه. وتُفسَّر قيم C_p الواقعة تحت الخط $C_p = p$ بأنها لا تعبر عن وجود انحياز، أي أن وقوعها تحت الخط يعود إلى خطأ المعاينة.

وعند استخدام المعيار C_p نهدف إلى تحديد مجموعات جزئية من المتغيرات X يكون من أجلها (١) قيمة C_p صغيرة و (٢) قيمة C_p قريبة من p. فللمجموعات من المتغيرات X التي تنتج قيما صغيرة لـ C_p متوسط مربعات خطأ صغير، ومع كون C_p قريبة من p أيضا يكون انحياز نموذج الانحدار صغيرا. وقد يحدث أحيانا أن ينطوي نموذج الانحدار الذي يستند إلى مجموعة جزئية من المتغيرات X موافقة لأصغر قيمة لـ C_p على انحياز كبير. وفي هذه الحالة فقد نفضل أحيانا نموذج انحدار مبني على مجموعة جزئية من المتغيرات X أكبر إلى حد ما، وقيمة C_p من أجلها أكبر بقليل، إلا أنه لاينطوي على مركبة انحياز كبيرة. ويتضمن المرجع [12.1] مناقشة مستفيضة لتطبيقات المعيار.

مثال. يتضمن الجدول (١٢-٣)، العمود (٦) قيم C_p لجميع نماذج الانحدار الممكنة في مثال وحدة الجراحة. وعلى سبيل المثال عندما يكون X_4 المتغير X الوحيد في نموذج الانحدار فإن قيمة C_p هي:

$$C_2 = \frac{SSE(X_4)}{MSE(X_1, X_2, X_3, X_4)} - [n - 2(2)] = \frac{1.8776}{.00224} - (54 - 4) = 788.2$$

والقيم C_p لجميع نماذج الانحدار الممكنة مرسومة في الشكل (١٢-٦). ونجد ثانية أن المجموعة الجزئية (X_1, X_2, X_3) مقترحة. وتحقق هذه المجموعة الجزئية أصغر قيمة لِـ C_p، وليس هناك مايشير إلى أي انحياز في نموذج الانحدار. وكما عرضنا سابقا فإن وقوع قيمة C_p لهذا النموذج، وهي $C_4 = 3.1$، تحت $p = 4$ هو نتيجة لتغير عشوائي في تقدير C_p.

ونلاحظ أن استخدام جميع المتغيرات X المرشحة (X_1, X_2, X_3, X_4) قد ينطوي على متوسط مربعات خطأ إجمالي أكبر. وكذلك فإن استخدام المجموعة الجزئية (X_2, X_3, X_4) بقيمة لِ C_p هي $C_4 = 161.7$، هو اختيار غير مقبول بسبب الانحياز الكبير للنموذج.

تعليقات

١- يتطلب الاستخدام المجدي للمعيار C_p تطويرا حذرا للمجموعات من $P - 1$ من المتغيرات X المرشحة، مع التعبير عن المتغيرات المستقلة بشكل مناسب (خطية، تربيعية، محوّلة) واستبعاد المتغيرات التي لافائدة منها بحيث يشكل $MSE(X_1,..., X_{P-1})$ تقديرا غير منحاز لتباين الخطأ σ^2.

٢- يؤكد المعيار C_p بشدة على توفيق نموذج المجموعة الجزئية للمشاهدات الـ n في العينة. ونفضل، أحيانا، تعديلا للمعيار C_p يؤكد على التنبؤ بمشاهدات جديدة.

٣- ولرؤية أن C_p كما عرفناها في (12.9) مقدِّر لِ Γ_p، نحتاج إلى الاستفادة من نتيجتين نعرضهما بصورة مبسطة. إذ يمكن، أولا، تبيان أن :

$$\sum_{i=1}^{n} \sigma^2\{\hat{Y}_i\} = p\sigma^2 \qquad (12.11)$$

وهكذا فإن الخطأ العشوائي الكلي للقيم التوفيقية $\hat{Y}_i$، وعددها n، يزداد بازدياد عدد المتغيرات في نموذج الانحدار.

وفضلا عن ذلك، يمكن تبيان أن:

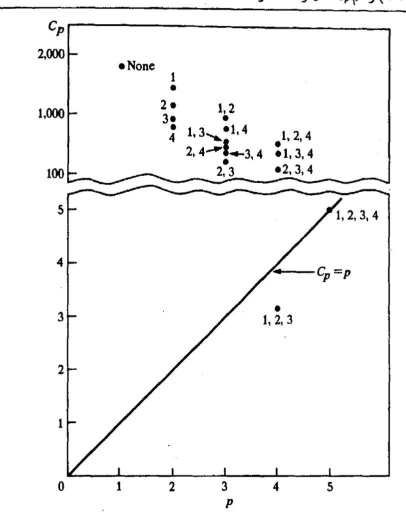
$$E\{SSE_p\} = \sum\left(E\{\hat{Y}_i\} - \mu_i\right)^2 + (n-p)\sigma^2 \qquad (12.12)$$

وهكذا يمكن التعبير عن Γ_p في (12.8) كما يلي:

$$\Gamma_p = \frac{1}{\sigma^2}\left[E\{SSE_p\} - (n-p)\sigma^2 + p\sigma^2\right] = \frac{E\{SSE_p\}}{\sigma^2} - (n-2p) \qquad (12.13)$$

وباستبدال المقدر SSE_p بِـ $E\{SSE_p\}$ واستخدام $MSE(X_1,..., X_{p-1})$ كمقدِّر لـ σ^2 نجد C_p كما عرفناها في (12.9).

شكل (١٢-٦) رسم C_p لمثال وحدة الجراحة



المعيار $PRESS_p$

يستند معيار الاختيار ***PRESS*** (مجموع مربعات التنبؤ) على رواسب الحذف d_i، المعرفة في (11.20):

$$d_i = Y_i - \hat{Y}_{i(i)} \tag{12.14}$$

حيث $\hat{Y}_{i(i)}$ هي القيمة التي تنبأنا بها للمشاهدة i عند توفيق دالة الانحدار بدون المشاهدة i. ولكل نموذج انحدار n من رواسب الحذف الموافقة له؛ والمعيار ***PRESS*** هو ببساطة مجموع مربعات رواسب الحذف هذه:

$$PRESS_p = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_{i(i)})^2 \qquad (12.15)$$

وتعتبر النماذج ذات القيم الصغيرة للمعيار **PRESS** نماذج مرشحة جيدة. وسبب ذلك هو أن راسب الحذف d_i يمثل خطأ التنبؤ عند توفيق نموذج انحدار بدون المشاهدة i ثم استخدامه للتنبؤ بـ $\hat{Y}_{i(i)}$ قيمة المشاهدة i. وتنطوي أخطاء التنبؤ الصغيرة على قيم صغيرة لـ d_i، أي على قيم صغيرة لـ d_i^2 وبالتالي على قيمة صغيرة لمجموع d_i^2. وهكذا فإن النماذج الموافقة لقيم صغيرة لـ **PRESS** ستشكل توفيقا جيدا بمعنى أن أخطاء التنبؤ فيها ستكون صغيرة.

ويمكن حساب قيم **PRESS** دون إعادة حسابات الانحدار مع تغير المشاهدة المحذوفة. وذلك باستخدام العبارة المكافئة لـ d_i كما وجدناها في (11.20a)

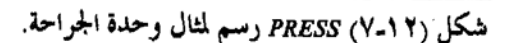
$$d_i = \frac{e_i}{1-h_{ii}}$$

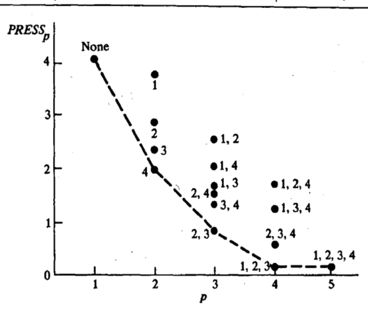
ويصبح $PRESS_p$:

$$PRESS_p = \sum_{i=1}^{n} \left(\frac{e_i}{1-h_{ii}} \right) \qquad (12.16)$$

حيث e_i الراسب العادي و h_{ii} قيمة العزم في (11.13) وكلاهما محسوب على أساس n مشاهدة.

مثال. يتضمن الجدول (١٢-٣) العمود (٧)، قيم الـ **SPRES** لجميع نماذج الانحدار الممكنة في مثال وحدة الجراحة. وقد رُسمت قيم الـ **PRESS** في الشكل (١٢-٧). وقد وصلنا القيم الدنيا للمعيار $(PRESS)_p$ بخطوط متقطعة. والرسالة التي يبلّغها الشكل (١٢-٧) مشابهة لتلك التي بلّغتها المعايير الأخرى. فالمجموعة الجزئية (X_1, X_2, X_3) مقترحة كمجموعة معقولة، إذ أن قيمة **PRESS** الموافقة لها هي القيمة الأصغر. وسينطوي استخدام جميع المتغيرات X المرشحة (X_1, X_2, X_3, X_4) على قيمة أكبر بقليل لـ **PRESS** ولو كنا نرغب في شمول النموذج للمتغير X_4 حيث $p = 4$ فستكون المجموعة الجزئية (X_2, X_3, X_4) هي المجموعة المقترحة إذ تنطوي على قيمة لـ **PRESS** تساوي 0.5487، وهي تتجاوز، باعتدال، القيمة 0.1405 لـ **PRESS** الخاصة بالمجموعة الجزئية (X_1, X_2, X_3).


شكل (١٢-٧) *PRESS* رسم لمثال وحدة الجراحة.



ملاحظة

يمكن أن تكون قيم المعيار مفيـدة أيضـا في عمليـة التحقـق مـن صلاحيـة نموذج كمـا سيجري إيضاحه في الفقرة (١٢-٦).

خوارزميات المجموعات الجزئية " الأفضل"

تمّ تطوير خوارزميات موفّرة للوقت وتحدد المجموعة الجزئيـة "الأفضـل"، وفقـا لمعيـار معين، دون الحاجة إلى توفيـق جميـع نمـاذج الانحـدار الممكنـة. ولاتحتـاج هـذه الخوارزميات، في الحقيقة، إلا لحساب كسر بسـيط مـن مجموعـة كـل نمـاذج الانحـدار الممكنة. وهكذا إذا كنا سنستخدم المعيار C_p، ونريد معرفة المجموعات الجزئية الخمـس "الأفضل" وفقا لهذا المعيـار، فـإن هـذه الخوارزميـات تبحـث عـن المجموعـات الجزئيـة الخمس من المتغيرات X التي تؤدي إلى القيم الخمس الأصغر لـ C_p مستخدمة جهـودا حسابية أقل بكثير من الجهود التي تتطلبها حسابات جميع المجموعـات الجزئيـة الممكنـة. وتسمى هذه الخوارزميات: خوارزميات المجموعات الجزئية "الأفضـل". ولاتقـدم هـذه

الخوارزميات المجموعات الجزئية الأفضل وفقا لمعيار معين فحسب، ولكنها تقدم، في الغالب أيضا عددا من المجموعات الجزئية "الجيدة" لكل عدد ممكن من المتغيرات X في النموذج، وبحيث تعطي الباحث معلومات إضافية مفيدة لاتخاذ قراره النهائي في اختيار المجموعة الجزئية من المتغيرات X التي سيستخدمها في نموذج الانحدار.

وعندما تكون جملة المتغيرات المستقلة المرشحة كبيرة جدا، لنقُل أكبر من 40 إلى 60، فحتى خوارزميات المجموعات الجزئية " الأفضل " قد تتطلب فترة طويلة من زمن الحاسوب. وتحت شروط كهذه قد نحتاج إلى استخدام أحد طرق الاختيار الآلية الموصوفة في الفقرة (١٢-٤)، للمساعدة في اختيار المتغيرات المستقلة.

مثال. في مثال وحدة الجراحة سيقدم استخدام إحدى خوارزميات المجموعات الجزئية "الأفضل" جزءا من المعلومات في الجدول (١٢-٣). وعند استخدام المعيار C_p ونريد التعرّف على "أفضل " ثلاث مجموعات جزئية فستحدد الخوارزمية المجموعات الجزئية (X_1, X_2, X_3)، (X_1, X_2, X_3, X_4)، و(X_2, X_3, X_4) كمجموعات جزئية تؤدي إلى أصغر ثلاث قيم لـِ C_p وبالإضافة إلى ذلك، يمكن أن تقدم الخوارزمية معلومات عن المجموعات الجزئية الثلاث "الجيدة" لكل مستوى من مستويات p.

بعض التعليقات الختامية

يقود أسلوب الاختيار الذي يعتمد على جميع الانحدارات الممكنة إلى تحديد عدد صغير من المجموعات الجزئية التي تُعتبر "جيدة" وفقا لمعيار معين. وبينما أشارت كل من المعايير الثلاثة، في مثال وحدة الجراحة، إلى المجموعات الجزئية " الأفضل " نفسها، فليست هذه هي الحال دائما، ولذلك فإنه من المستحسن أحيانا اعتبار أكثر من معيار عند تثمين مجموعات جزئية ممكنة من المتغيرات X.

وحالما يتعرف الباحث على عدد قليل من المجموعات الجزئية " الجيدة" بغية فحصها بصورة مركّزة ومستفيضة، فإنه لابد من القيام باختيار نهائي لمتغيرات النموذج. ومما يساعد في هذا الاختيار، كما يشير استراتيجنا لبناء نموذج في الشكل (١٢-١)، القيام بتحاليل للراسب، وفحص المشاهدة المؤثرة، وتشخيصات أخرى لكل

من النماذج المتنافسة، ومعرفة الباحث بالظاهرة المدروسة، ثم التثبت في الختام من صحة النموذج.

(٤-١٢) انحدار الخطوة فخطوة إلى الأمام، أساليب بحث آلية أخرى، واستخدام انحدار الحافة لتخفيض عدد المتغيرات

في تلك الحالات حيث يتفق أن تتضمن جملة المتغيرات X المرشحة من 40 إلى 60 متغيرا أو حتى أكثر من ذلك، فقد لايكون استخدام خوارزمية المجموعات الجزئية "الأفضل" أمرا عمليا. وقد يكون من المفيد في تلك الحالات اللجوء إلى طريقة بحث آلية تكشف بصورة تتابعية عن المجموعة الجزئية من المتغيرات X التي سيشملها النموذج. ربما كانت طريقة الانحدار خطوة فخطوة إلى الأمام هي أوسع طرق البحث الآلي استخداما. وقد طُوِّرت هذه الطريقة للتوفير في الجهود الحسابية اللازمة للوصول إلى المجموعة "الأفضل" من المتغيرات المستقلة، وذلك بالمقارنة مع أسلوب جميع الانحدارات الممكنة، وفي جوهرها فإن هذه الطريقة في البحث تكشف عن متتابعة من نماذج الانحدار، فتضيف عند كل خطوة، أو تحذف، متغيرا من المتغيرات X. ويمكن التعبير عن قاعدة إضافة أو حذف متغير X بدلالة تخفيض مجموع مربعات الخطأ، أو بدلالة معامل الارتباط الجزئي، أو بدلالة الإحصاءة F^* على التكافؤ.

والفرق الأساسي بين طرق البحث الآلية وأسلوب جميع الانحدارات الممكنة هو أن طرق البحث الآلية تنتهي بتحديد نموذج انحدار بمفرده "كأفضل" نموذج، بينما يمكن لأسلوب جميع الانحدارات الممكنة، على الوجه الآخر، أن يرشح عدة نماذج انحدار كنماذج " جيدة" في الاعتبار النهائي.

انحدار الخطوة فخطوة إلى الأمام

سنصف خوارزمية البحث الخاصة بانحداز الخطوة فخطوة إلى الأمام بدلالة الإحصاءة F^* لاختبار F الجزئي.

١- يقوم روتين انحدار الخطوة فخطوة أولا بتوفيق نموذج انحدار خطي بسيط

لكل من المتغيرات X المرشحة وعدتها $P - 1$ ولكل نموذج انحدار خطي بسيط نحصل على الإحصاءة F^* في (3.59) لاختبار ما إذا كان الميل مساويا للصفر أم لا:

$$F_k^* = \frac{MSR(X_k)}{MSE(X_k)} \tag{12.17}$$

لنتذكر أن $MSR(X_k) = SSR(X_k)$ يقيس الانخفاض في التغير الكلي لـ Y المصاحب لاستخدام المتغير X_k. والمتغير X الذي يؤدي إلى أعلى قيمة لـ F^* هو المتغير المرشح كأول إضافة. وإذا تجاوزت قيمة F^* مستوى محددا سلفا يضاف المتغير X، وفيما عدا ذلك ينتهي البرنامج معتبرا أنه لايوجد أي متغير X مفيد بما يكفي لدخول نموذج الانحدار.

٢- افترض أن المتغير X_7 هو المتغير الذي دخل النموذج في الخطوة الأولى. فيقوم روتين انحدار الخطوة فخطوة بتوفيق جميع نماذج الانحدار المتضمنة لمتغيرين مستقلين أحدهما X_7 . ولكل نموذج انحدار كهذا نحصل على إحصاءة الاختبار الجزئي (8.22):

$$F_k^* = \frac{MSR(X_k \mid X_7)}{MSE(X_7, X_k)} = \left(\frac{b_k}{s\{b_k\}}\right)^2 \tag{12.18}$$

وهي إحصاءة لاختبار ما إذا كان $\beta_k = 0$ عندما يشكل X_7 و X_k جميع متغيرات النموذج. والمتغير X بأعلى قيمة لِـ F^* هو المرشح للإضافة في الخطوة الثانية. فإذا تجاوزت قيمة F^* مستوى محددا سلفا يُضاف المتغير الثاني ذاك، وفيما عدا ذلك ينتهي البرنامج.

٣- افترض أن X_3 قد أضيف في الخطوة الثانية. فيفحص روتين انحدار الخطوة فخطوة الآن ما إذا كان ينبغي حذف أي من المتغيرات X الموجودة في النموذج، وفي توضيحنا هنا، يوجد متغير X آخر في النموذج هو المتغير X_7 بمفرده، وبالتالي نحصل على إحصاءة اختبار F جزئي واحدة فقط هي:

$$F_7^* = \frac{MSR(X_7 \mid X_3)}{MSE(X_3, X_7)} \tag{12.19}$$

وفي مراحل لاحقة سيكون هناك عدد من هذه الإحصاءات F^* واحدة لكل من المتغيرات في النموذج إلى جانب آخر متغير أُضيف. والمتغير الذي تكون قيمة F^* من

أجله هي القيمة الأصغر هو المتغير المرشح للحذف. وإذا كانت هذه القيمــة الصغـرى لـِ F^* أقل من قيمة محددة سلفا يجري إسقاط المتغير من النموذج وفيما عدا ذلك نحتفظ به.

٤ـ لنفترض أننـا احتفظنـا بـ X_7، فـالنموذج يتضمـن الآن كـلا مـن X_3 و X_7. ويفحص روتين انحدار الخطوة فخطوة الآن المرشح التالي من بين المتغيرات X للإضافـة إلى النموذج، ثم يفحص بعدئذ ما إذا كان ينبغي شطب أي من المتغيرات الموجودة من حينها في النموذج. وهكذا حتى نصل إلى مرحلة لايمكن فيها إضافة أوحـذف أي مـن المتغيرات X، وعندها تنتهي عملية البحث.

وتنبغي ملاحظة أن خوارزمية انحدار الخطوة فخطوة تسمح بإدخال متغـير X إلى النموذج في مرحلة مبكـرة لتحذفـه فيمـا بعـد إذا لم يعـد مفيـدا مـع وجـود متغـيرات أضيفت في مراحل لاحقة.

مثال. يبين الشكل (١٢ـ٨) مطبوعة حاسب تمّ الحصول عليها عند تطبيق روتين معين للانحدار خطوة فخطوة (***BMDP2R***، مرجع [12.2]) على مثال وحدة الجراحـة. وقـد حُددت أقل قيمة مقبولة لِـ F من أجل إضافة متغير وأعلى قيمة مقبولة لِـ F مـن أجـل حذف متغير على أنها 4.0 و3.9، على الـترتيب، وذلـك كمـا هـو موضـح في الزاويـة العليا اليمنى من الشكل (١٢ـ٨). وبما أن درجات الحرية المصاحبة لِـ MSE تتغير وفقا لعدد المتغيرات X في النمـوذج، وأننـا نقـوم باختبـارات متكـررة مسـتخدمين البيانـات نفسها، فليس لقيمتيَ F المثبتتين والخاصتين بإضافة متغـير أو حذفـه أي معنـى احتمـالي دقيق. ونلاحظ، على أي حال، أن $F(0.95; 1,50) = 4.03$ أي أن القيمتين المحددتين لِـ F وهما 4.0 و3.9 تقابلان، علـى وجـه التقريـب، مسـتوى معنويـة 0.05 لأي اختبـار بمفرده يستند إلى 50 درجة حرية تقريبا.

وحد التساهل المقبول الأدنـى 0.01 المبـين في الزاويـة العليـا اليمنـى مـن الشـكل (١٢ـ٨) هو تحديد يهدف إلى الحماية من دخول متغير يرتبط ارتباطا عاليـا بالمتغـيرات X الأخرى الموجودة من حينها في النموذج. وكما شرحنا في الفقرة (١١ــ٦)، يُعـرّف التساهل بأنه $1-R_k^2$ حيث R_k^2 معامل التحديد المتعدد عند انحدار X على المتغيرات X

الأخرى في النموذج. وتحديد التساهل بأنه 0.01 في الشكل (١٢ــ٨) يؤدي إلى عدم إضافة أي متغير إلى النموذج إذا تجاوز معامل تحديده المتعدد مع المتغيرات X الأخرى في النموذج المقدار $0.99 = 0.1 - 1$، أو عندما يسبب تجاوز R_k^2 الخاصة بأي متغير في النموذج للقيمة 0.99.

وسنتابع الآن خطوات العمل بالكامل:

١ــ في الخطوة 0، لايوجد أي متغيرX في النموذج، والنموذج الذي سنقوم بتوفيقه هو $Y_i = \beta_0 + \varepsilon_i$. والراسب أو مجموع مربعات الخطأ المبين في جدول التحاين في الشكل (١٢-٨) للخطوة 0 هو بالتالي: $\Sigma(Y_i - \bar{Y})^2 = SSTO = 3.9728$. وتُحسب لكل متغير X مُرشَّح، الإحصاءة F^* في (12.17). وهذه القيم F^* مبينة في الشكل (١٢ــ٨) تحت عنوان "متغيرات ليست في المعادلة (*Variables not in equation*) وتسمى قيم F^* – للدخول *(F to enter)*. ونرى أن F_4^* هي القيمة الأكبر:

$$F_4^* = \frac{MSR(X_4)}{MSE(X_4)} = \frac{2.095140}{.03610831} = 58.02$$

وبما أن هذه القيمة تتجاوز قيمة "F – للدخول" *(F to enter)* الدنيا وهي 4.0 فيُضاف X_4 إلى النموذج.

والعمود تحت عنوان "مستوى" *(level)* يشير إلى نوع من الاختيار الذي يسمح للمستخدم بمنح أولويات مختلفة للمتغيرات X المرشحة. وفي المثال الحالي، نلاحظ أن لجميع المتغيرات X الأولوية نفسها.

٢ـ في هذه المرحلة، نستكمل الخطوة ١. نموذج الانحدار الراهن يتضمن X_4 وقد أُعطيت معاملات الانحدار المقدَّرة، وجدول تحليل التباين، بالإضافة إلى معلومات مختارة عن النموذج الراهن.

بعدها يتم توفيق جميع نماذج الانحدار التي تتضمن X_4 بالإضافة إلى متغير مستقل آخر، وتُحسب الإحصاءات F^* وهي الآن:

$$F_k^* = \frac{MSR(X_k \mid X_4)}{MSE(X_4, X_k)}$$

وهـذه الإحصـاءات مبينـة في الخطـوة ١: تحـت عنـوان "متغـيرات ليسـت في المعادلـة" وللمتغير X_3 أعلى قيمة لِ F^*، وهي قيمة تتجاوز 4.0 فالمتغير X_3 يدخل الآن النموذج.

٣- وتلخص الخطوة ٢ في الشكل (١٢-٨) الحالة عنـد هـذه النقطـة مـن مسـار العملية. فالنموذج يتضمن الآن X_3 و X_4 وقد أُعطيت معلومات حول هـذا النمـوذج. والآن نقوم باختبار ما إذا كـان ينبغـي حـذف X_4 أم لا. والإحصـاءة F^* مبينـة تحـت عنوان "متغيرات في المعادلة" وقد سُمِّيت "F ـ للحذف" (F to remove):

$$F_4^* = \frac{MSR(X_4 \mid X_3)}{MSE(X_3, X_4)} = 39.72$$

وبما أن هذه القيمة لِF^* تتجاوز قيمة"F ـ للحذف" العظمى وهي 3.9 فلا نحذف X_4.

٤- نقوم الآن بتوفيق جميع نماذج الانحدار التي تتضمن X_3 وX_4 وإحدى المتغـيرات المرشحة X. والإحصاءات F^* المناسبة الآن هي:

$$F_k^* = \frac{MSR(X_k \mid X_3, X_4)}{MSE(X_3, X_4, X_k)}$$

وهذه الإحصاءات مبينة في الخطوة رقم 2 تحـت عنـوان "متغـيرات ليسـت في المعادلـة" وللمتغير X_2 أعلى قيمة لِ F^* وهي قيمة تتجاوز 4.0، فالمتغير X_2 يدخل الآن النموذج.

٥- وتلخص الخطوة رقم 3 في الشكل (١٢-٨) الحالة عند هذه النقطـة. ويشـمل النموذج الآن X_2، X_3 و X_4. بعدها يجري اختبار ما إذا كان ينبغـي حـذف X_3 أو X_4. والاحصاءات F^* لحذف متغير مبينة تحت العنوان " متغـيرات في المعادلـة " في الخطـوة رقم 3. F_4^* هو الأصغر:

$$F_4^* = \frac{MSR(X_4 \mid X_2, X_3)}{MSE(X_2, X_3, X_4)} = 29.86$$

وبما أن قيمته تتجاوز 3.9 فلا نحذف X_4 من النموذج.

٦- عند هذه النقطة لايبقى مـن جملـة المتغـيرات المرشـحة إلا X_1. وقيمـة F^* للدخـول الخاصة بـه تتجـاوز 4.0 (انظـر "المتغـيرات ليسـت في المعادلـة" تحـت الخطـوة رقـم 3) وهكذا يدخل X_1 إلى النموذج.

شكل (١٢-٨) انحدار الخطوة فخطوة إلى الأمام لمثال وحدة الجراحة (*BMDP2R* مرجع [12.2])

```
STEPPING ALGORITHM. . . . . . . . . . . . . . .F        MINIMUM ACCEPTABLE F TO ENTER . . . . . . . . .   4.000,   4.000
MAXIMUM NUMBER OF STEPS . . . . . . . . . .    10       MAXIMUM ACCEPTABLE F TO REMOVE. . . . . . . . .   3.900,   3.900
DEPENDENT VARIABLE. . . . . . . . . . . . .     5 y     MINIMUM ACCEPTABLE TOLERANCE. . . . . . . . . . 0.01000
                                                        SUBSCRIPTS OF THE INDEPENDENT VARIABLES . . . .   1   2   3   4
STEP NO.   0
---------------

STD. ERROR OF EST.     0.2736                  ANALYSIS OF VARIANCE
                                                           SUM OF SQUARES     DF   MEAN SQUARE
                                                RESIDUAL    3.9727724         53  0.7495797E-01 <-- SSTO/(n-1)
                                                 SSTO -->
             VARIABLES IN EQUATION FOR y                    .         VARIABLES NOT IN EQUATION

                          STD. ERROR  STD REG                 F               .               PARTIAL               F
   VARIABLE    COEFFICIENT  OF COEFF    COEFF   TOLERANCE  TO REMOVE  LEVEL.   VARIABLE      CORR.   TOLERANCE  TO ENTER   LEVEL
 (Y-INTERCEPT     2.20614 )                                                   .
                  ^ Y-bar                                                     . x1      1  0.34640  1.00000      7.09       1
                                                                              . x2      2  0.59289  1.00000     28.19       1
                                                                              . x3      3  0.66512  1.00000     41.25       1
                                                                              . x4      4  0.72621  1.00000     58.02       1
STEP NO.   1                                                                               ^ r_Yk                ^ F*_k
---------------
VARIABLE ENTERED    4 x4                       ANALYSIS OF VARIANCE                              MSR(X4)
                                  SSR(X4) -->              SUM OF SQUARES     DF   MEAN SQUARE          F RATIO
MULTIPLE R           0.7262 <-- R               REGRESSION  2.0951402          1   2.095140 <--           58.02
MULTIPLE R-SQUARE    0.5274 <-- R^2             RESIDUAL    1.8776320         52  0.3610831E-01           ^ F*
ADJUSTED R-SQUARE    0.5181 <-- R^2_a                       ^ SSE(X4)                ^ MSE(X4)

STD. ERROR OF EST.   0.1900 <-- sqrt(MSE)
             VARIABLES IN EQUATION FOR y                    .         VARIABLES NOT IN EQUATION

                          STD. ERROR  STD REG                 F               .               PARTIAL               F
   VARIABLE    COEFFICIENT  OF COEFF    COEFF   TOLERANCE  TO REMOVE  LEVEL.   VARIABLE      CORR.   TOLERANCE  TO ENTER   LEVEL
 (Y-INTERCEPT     1.69639 )                                                   .
 x4          4    0.18575     0.0244    0.726    1.00000     58.02      1 . x1      1 -0.03104  0.74758      0.05       1
                  ^ b_k       ^ s(b_k)  ^ b'_k               ^ F*_k           . x2      2  0.50849  0.86382     17.78       1
                                                                              . x3      3  0.58031  0.82659     25.89       1
STEP NO.   2                                                                               ^ r_Yk.4 ^ 1-r^2_4k   ^ F*_k
---------------
VARIABLE ENTERED    3 x3                       ANALYSIS OF VARIANCE
                                                           SUM OF SQUARES     DF   MEAN SQUARE          F RATIO
MULTIPLE R           0.8286                     REGRESSION  2.7274444          2   1.363722              55.85
MULTIPLE R-SQUARE    0.6865                     RESIDUAL    1.2453278         51  0.2441819E-01
ADJUSTED R-SQUARE    0.6742

STD. ERROR OF EST.   0.1563

             VARIABLES IN EQUATION FOR y                    .         VARIABLES NOT IN EQUATION

                          STD. ERROR  STD REG                 F               .               PARTIAL               F
   VARIABLE    COEFFICIENT  OF COEFF    COEFF   TOLERANCE  TO REMOVE  LEVEL.   VARIABLE      CORR.   TOLERANCE  TO ENTER   LEVEL
 (Y-INTERCEPT     1.38878 )                                                   .
 x3          3    0.00565     0.0011    0.439    0.82659     25.89      1 . x1      1  0.32276  0.59179      5.81       1
 x4          4    0.13901     0.0221    0.543    0.82659     39.72      1 . x2      2  0.79148  0.82580     83.85       1
                                                                                                 ^ 1-R^2_k
STEP NO.   3
---------------
VARIABLE ENTERED    2 x2                       ANALYSIS OF VARIANCE
                                                           SUM OF SQUARES     DF   MEAN SQUARE          F RATIO
MULTIPLE R           0.9396                     REGRESSION  3.5075624          3   1.169187             125.66
MULTIPLE R-SQUARE    0.8829                     RESIDUAL    0.46520978        50  0.9304196E-02
ADJUSTED R-SQUARE    0.8759

STD. ERROR OF EST.   0.0965
             VARIABLES IN EQUATION FOR y                    .         VARIABLES NOT IN EQUATION

                          STD. ERROR  STD REG                 F               .               PARTIAL               F
   VARIABLE    COEFFICIENT  OF COEFF    COEFF   TOLERANCE  TO REMOVE  LEVEL.   VARIABLE      CORR.   TOLERANCE  TO ENTER   LEVEL
 (Y-INTERCEPT     0.94226 )                                                   .
 x2          2 0.78987E-02 0.8626E-03   0.488    0.82580     83.85      1 . x1      1  0.87409  0.55586    158.66       1
 x3          3 0.69997E-02 0.7013E-03   0.543    0.79021     99.63      1 .
 x4          4    0.08185     0.0150    0.320    0.68298     29.86      1 .
STEP NO.   4
---------------
VARIABLE ENTERED    1 x1                       ANALYSIS OF VARIANCE
                                                           SUM OF SQUARES     DF   MEAN SQUARE          F RATIO
MULTIPLE R           0.9861                     REGRESSION  3.8630006          4  0.9657502             431.09
MULTIPLE R-SQUARE    0.9724                     RESIDUAL    0.10977174        49  0.2240240E-02
ADJUSTED R-SQUARE    0.9701

STD. ERROR OF EST.   0.0473
             VARIABLES IN EQUATION FOR y                    .         VARIABLES NOT IN EQUATION

                          STD. ERROR  STD REG                 F               .               PARTIAL               F
   VARIABLE    COEFFICIENT  OF COEFF    COEFF   TOLERANCE  TO REMOVE  LEVEL.   VARIABLE      CORR.   TOLERANCE  TO ENTER   LEVEL
 (Y-INTERCEPT     0.48876 )                                                   .
 x1          1    0.06852     0.0054    0.401    0.55586    158.66      1 .
 x2          2 0.92541E-02 0.4367E-03   0.571    0.77567    448.98      1 .
 x3          3 0.94745E-02 0.3963E-03   0.736    0.59594    571.70      1 .
 x4          4    0.00193     0.0097    0.008    0.39134      0.04      1 .
STEP NO.   5
---------------
VARIABLE REMOVED    4 x4                       ANALYSIS OF VARIANCE
                                                           SUM OF SQUARES     DF   MEAN SQUARE          F RATIO
MULTIPLE R           0.9861                     REGRESSION  3.8629124          3   1.287637             586.04
MULTIPLE R-SQUARE    0.9723                     RESIDUAL    0.10985984        50  0.2197197E-02
ADJUSTED R-SQUARE    0.9707

STD. ERROR OF EST.   0.0469

             VARIABLES IN EQUATION FOR y                    .         VARIABLES NOT IN EQUATION

                          STD. ERROR  STD REG                 F               .               PARTIAL               F
   VARIABLE    COEFFICIENT  OF COEFF    COEFF   TOLERANCE  TO REMOVE  LEVEL.   VARIABLE      CORR.   TOLERANCE  TO ENTER   LEVEL
 (Y-INTERCEPT     0.48362 )                                                   .
 x1          1    0.06923     0.0041    0.405    0.97011    288.16      1 . x4      4  0.02832  0.39134      0.04       1
 x2          2 0.92945E-02 0.3825E-03   0.574    0.99177    590.44      1 .
 x3          3 0.95236E-02 0.3064E-03   0.739    0.97751    966.06      1 .

***** F LEVELS(  4.000,  3.900) OR TOLERANCE INSUFFICIENT FOR FURTHER STEPPING
```

٧ـ وتلخص الخطوة رقم 4 في الشكل (١٢ـ ٨) إضافة المتغير X_1 إلى نموذج يتضمن المتغيرات X_2، X_3 وX_4. وبعدها يجري اختبار ما إذا كان ينبغي حذف X_2 أو X_3 أو X_4. والاحصاءات *F مبينة تحت العنوان "متغيرات في المعادلة" في الخطوة 4. لاحظ أن:

$$F_4^* = \frac{MSR(X_4 \mid X_1, X_2, X_3)}{MSE(X_1, X_2, X_3, X_4)} = 0.04$$

هو الأصغر، وقيمته أقل من 3.9 وبالتالي نلغي X_4.

٨ـ وتلخص الخطوة رقم 5 إلغاء X_4 من النموذج. وبما أن المتغير المرشح الوحيد الباقي هو X_4، وقد ألغي لتوه من النموذج، فلا يمكن إدخاله إلى النموذج الآن، وبالتالي تدرس الخوارزمية الآن قيم "F ـ للحذف" في الخطوة رقم 5 وهي تشير إلى أن F_1^* هي الأصغر. وبما أن هذه القيمة لاتتجاوز 3.9 فلا يُحذف X_1 من النموذج وتنتهي عملية البحث.

وهكذا فإن خوارزمية البحث خطوة فخطوة تحدد (X_1 , X_2 , X_3) "كأفضل" مجموعة جزئية من المتغيرات X، وهي نتيجة يتفق أن تكون منسجمة مع تحليلاتنا السابقة القائمة على أسلوب جميع الانحدارات الممكنة.

تعليقات

١ـ في مثال وحدة الجراحة، كانت متطلبات حد التساهل محققة دائما، وبالتالي لم نستثن أي متغير من النموذج كنتيجة لارتباط مرتفع بينه وبين المتغيرات X الأخرى في النموذج.

٢ـ يمكن أن تتغير القواعد الخاصة بدخول متغيرات وحذفها والموضحة في المثال. وعلى سبيل المثال، يمكن استخدام قيم لـ "F ـ للدخول" و"F ـ للحذف" وذلك وفقا لعدد درجات الحرية المصاحب لِ MSE في الإحصاءة *F. إلا أن مثل هذا التحسين لا يُستخدم، في الغالب، وتُستخدم بدلا من ذلك قيم مثبتة، باعتبار أن الاختبارات المتكررة في طريقة البحث لا تسمح بتفسيرات احتمالية دقيقة.

٣- لاحاجة لاختيار القيمتين الحديتين لِ F من أجل إضافة متغير أو حذفه بدلالة مستويات معنوية تقريبية، بل يمكن تحديدها بصورة وصفية بدلالة التخفيض في الخطأ. وعلى سبيل المثال، يمكن تحديد 2.0 كقيمة F من أجل إضافة متغير مع التصور بأنه حالما نضيف المتغير فإن التخفيض الهامشي للخطأ المصاحب لهذه الإضافة ينبغي أن لايقل عن ضعف متوسط مربعات الخطأ الباقي.

٤- يمثل اختيار قيمتيّ "F - للدخول" و"F - للحذف" في الأساس نوعا من الموازنة بين نزعتين متضادتين. وقد بينت دراسات محاكاة أنه من أجل مجموعات كبيرة من المتغيرات المستقلة غير المرتبطة والتي تم توليدها بحيث لاتكون مرتبطة بالمتغير التابع، يُنتج استخدام قيم صغيرة أو صغيرة باعتدال لقيم "F - للدخول" طريقة منفتحة جدا، أي طريقة تسمح بدخول أكثر مما ينبغي من المتغيرات المستقلة إلى النموذج. وعلى الوجه الآخر، فإن النماذج الناتجة عن طريقة اختيار آلية بقيم كبيرة لـِ"F - للدخول" هي في الغالب نماذج مختزلة أكثر مما ينبغي، مما يُنتج فرط تقدير بالزيادة لـِ σ^2 وتكون الطريقة محافظة جدا، (انظر المرجعين [12.3] و[12.4]).

٥- يجب أن لاتكون القيمة الدنيا المقبولة لـ"F - للدخول" أقل أبدا من القيمة العظمى المقبولة لِ "F - للحذف" ؛ وإلا فيمكن الوقوع في حلقة مفرغة حيث يدخل متغير ثم يُلغى بصورة مستمرة.

٦- لايعكس ترتيب دخول المتغيرات إلى النموذج أهميتها النسبية. ففي مثال وحدة الجراحة، على سبيل المثال، كان X_4 أول متغير دخل النموذج، مع أنه في النهاية أُلغي.

٧- يطبع روتين انحدار الخطوة فخطوة الذي استخدمناه معاملات الارتباط الجزئي في كل مرحلة. وبصورة مكافئة، يمكن استخدام هذه القيم لغربلة المتغيرات X بدلا من القيم F^*، وفي الحقيقة تستخدم بعض الروتينات، في الواقع، معاملات الارتباط الجزئي للغربلة.

٨- وأحد المآخذ على طريقة بحث انحدار الخطوة فخطوة إلى الأمام هو أنها تفترض وجود مجموعة جزئية بمفردها كمجموعة "أفضل" من المتغيرات X وتهدف إلى

تحديدها. وكما ذكرنا سابقا، ليس هناك، في الغالب، مجموعة جزئية وحيدة يمكن اعتبارها المجموعة الجزئية "الأفضل". والمأخذ الآخر على روتين انحدار الخطوة فخطوة إلى الأمام هو أنها تصل أحيانا إلى مجموعة جزئية "أفضل" غير منطقية، وذلك عندما يكون الارتباط عاليا جدا بين المتغيرات X.

طرق بحث آلية أخرى

هناك عدد من طرق البحث الآلية التي اقتُرحت لإيجاد مجموعة جزئية " أفضل" من المتغيرات المستقلة. ونذكر هنا اثنتين منها. إلا أن أيا من الطريقتين لم تكتسب القبول الذي اكتسبته طريقة بحث الخطوة فخطوة إلى الأمام.

الاختيار بالإضافة. طريقة البحث هذه هي نسخة مبسطة من انحدار الخطوة فخطوة إلى الأمام. إذ تلغي اختبار ما إذا كان ينبغي حذف المتغير من النموذج بعد أن يتم دخوله إليه.

الاختيار بالحذف. وطريقة البحث هذه معاكسة تماما لطريقة الاختيار بالإضافة. إذ تبدأ بنموذج يتضمن جميع المتغيرات X المرشحة وتحدد المتغير الموافق لأصغر قيمة لِ F^* مثلا القيمة لِ X_1 هي:

$$F_1^* = \frac{MSR(X_1 \mid X_2,\ldots, X_{p-1})}{MSE(X_1,\ldots,X_{p-1})} \tag{12.20}$$

وإذا كانت القيمة الدنيا F_K^* لِ F^* أقل من قيمة محددة سلفا فيلغى المتغير X ويجري عندئذ توفيق النموذج المتضمن لمتغيرات التنبؤ الـ $p - 2$ الباقية، ثم يتم تحديد المرشح التالي للحذف. وتستمر هذه العملية حتى يصبح إلغاء المزيد من المتغيرات X غير ممكن. ويمكن تكييف هذه الطريقة أيضا بحيث يطرأ عليها تعديل "خطوة فخطوة" يسمح لمتغيرات كانت قد حذفت سابقا بأن تضاف في مرحلة لاحقة، ويدعى هذا التعديل طريقة انحدار الخطوة فخطوة إلى الوراء.

ملاحظة

يجادل بعض الإحصائيين لصالح البحث خطوة فخطوة إلى الوراء بالمقارنة مع البحث خطوة فخطوة إلى الأمام وذلك عندما يكون عدد المتغيرات في جملة المتغيرات X

المرشحة صغيرا أو معتدلا (انظر المرجع [12.5]) وتستند حجتهم في المقام الأول على حالات يكون من المفيد فيها كخطوة أولى النظر إلى كل متغير مستقل في دالة الانحدار بعد تعديله من أجل جميع المتغيرات المستقلة الأخرى في جملة المتغيرات X المرشحة.

خيارات حاسوبية وتحسينات

ركّزت مناقشتنا لطرق الاختيار الآلي الرئيسة (لأفضل) مجموعة جزئية من المتغيرات X على المسائل الفكرية الأساسية وليس على الخيارات والتغيرات والتحسينات المتوافرة في حزم حاسوبية بعينها. ومن المهم جدا أن نفهم تماما خصائص معينة للحزمة التي سنستخدمها، بحيث يمكن استخدام الحزمة بذكاء. ففي بعض الحزم يوجد خيار لنماذج انحدار عبر الأصل. وتسمح بعض الحزم بإدخال المتغيرات إلى النموذج أزواجا واختبارها أزواجا، أو وفق تجميعات أخرى، بدلا من إدخالها واختبارها فرادى، وذلك للتوفير في زمن الحسابات أو لأسباب أخرى. وبعض الحزم تقوم، حال التعرّف على نموذج الانحدار الأفضل، بتوفيق جميع نماذج الانحدار الممكنة التي تتضمن العدد نفسه من المتغيرات وتطوّر معلومات عن كل نموذج بحيث يمكن للمستخدم أن يحدد اختياره النهائي. وهناك اختيارات في بعض برامج الخطوة فخطوة لقسر متغيرات على نموذج الانحدار؛ ومثل هذه المتغيرات لاتُحذف حتى لو أصبحت قيم F^* الخاصة بها منخفضة جدا.

ويخدم تنوع هذه الاختيارات والخصائص للتأكيد على نقطة ذكرناها سابقا: لاتوجد طريقة وحيدة للبحث عن مجموعات جزئية " جيدة "من المتغيرات X ولابد لعناصر ذاتية أن تلعب دورا مهما في عملية البحث.

اختيار متغيرات باستخدام انحدار الحافة

ناقشنا في الفقرة (١١-٧) استخدام انحدار الحافة للمساعدة في التغلب على مشاكل تتصل بالخطية المتعددة بين المتغيرات X ويمكن أيضا استخدام أثر الحافة المذكور هناك (شكل (١١-٨)، صفحة ٥٤١) لتحديد المتغيرات التي يمكن حذفها من نموذج الانحدار. وقد اقتُرح إلغاء المتغيرات، التي يكون أثر الحافة الخاص بها غير مستقر، ومعاملها ينحو إلى القيمة صفر. وينبغي أيضا حذف المتغيرات التي يكون أثر

الحافةمن أجلها مستقرا ولكن عند قيمة صغيرة جدا. وأخيرا، ينبغي اعتبار المتغيرات ذات آثار الحافة غير المستقرة والتي لاتنحو في اتجاه الصفر كمتغيرات مرشحة للحذف.

(١٢-٥) تحسين النموذج واختياره

تُنتج غربلة المتغيرات، بعملية اختيار حاسوبية أو غيرها، في العادة، عددا صغيرا من النماذج المرشحة. وتحتاج هذه النماذج عندئذ إلى مزيد من الدراسة لمصداقيتها باستخدام طرق التشخيص في الفصلين الرابع والحادي عشر. وغالبا مايعتمد الاختيار النهائي لنموذج الانحدار اعتمادا كبيرا على نتائج التشخيص هذه. وعلى سبيل المثال، قد يتأثر توفيق نموذج تأثرا كبيرا بمشاهدة واحدة في حين لايتأثر نموذج آخر. ومن جديد، قد يُظهر أحد النماذج التوفيقية ارتباطات بين حدود الخطأ في حين لايُظهر نموذج آخر مثل ذلك.

وعندما تتوافر مشاهدات مكررة، يمكن القيام باختبارات رسمية لنقص التوفيق. وفي جميع الأحوال، يمكن استخدام تشكيلة من التحليلات ورسومات الراسب لتحديد أي نقص في التوفيق، أو مشاهدات قاصية، أو مشاهدات مؤثرة. وعندما تستثني المجموعة الأصلية من المتغيرات X المرشحة، وعدّتها $P - 1$، الحدود الجدائية والحدود الأسية في المتغيرات المستقلة، بغية حفظ مسألة الاختيار ضمن حدود منطقية، فقد يكون من المفيد القيام برسومات راسب مقابل مثل هذه المتغيرات" الغائبة"، أو زيادة كل من المجموعات الجزئية " الجيدة " من المتغيرات المستقلة بإضافة حدود جدائية و/أو حدود قوى، وذلك للتعرف على طرق يمكن أن تؤدي إلى مزيد من التحسين في توفيق النموذج.

وعند استخدام طريقة اختيار آلية وتحديد نموذج بمفرده بأنه النموذج "الأفضل"، فينبغي أيضا استطلاع نماذج أخرى. وإحدى الإجراءات الممكنة هي استخدام عدد المتغيرات المستقلة في النموذج الموسوم بأنه "الأفضل" كتقدير لعدد المتغيرات المستقلة التي نحتاجها في نموذج الانحدار. وعندئذ يستطلع الباحث ويحدد نماذج مرشحة أخرى تتضمن، على وجه التقريب، العدد نفسه من المتغيرات المستقلة الذي حددته الطريقة الآلية.

وسنوضح الآن تحسين النموذج وطور الاختيار النهائي في عملية بناء نمـوذج مـن خلال مثال وحدة الجراحة.

مثال

بما أننا بدأنا بجملة صغيرة من المتغيرات المستقلة المرشحة في مثال وحدة الجراحـة، فمن الممكن توليد جميع نمـاذج الانحـدار الممكنـة مـن المرتبـة الأولى، وذلـك مـن أجـل المتغيرات المستقلة الأربعة في الجملة. ولنتذكر أن جميـع المعايـير قـد اقـترحت النمـوذج المتضمن لـ X_1، X_2 وX_3 كأفضل نموذج. وبالتالي فسيركز اختيار وتحسين النمودج هنـا على مزيد مـن الدراسـة للانحنـاء وتأثيرات التفـاعل، والخطيـة المتعـددة، والمشـاهدات المؤثـرة، لنمـوذج الانحـدار المتضمـن لِـ X_1، X_2 و X_3، وذلـك باسـتخدام الرواسـب والتشخيصات الأخرى.

ولمزيد من الفحص لتأثيرات التفاعل، فقد تم توفيق نموذج انحدار يتضمن الحـدود من المرتبة الأولى في X_1، X_2 و X_3 وجميع حدود التفاعل الثنائية ورُسمـت رواسـب هـذا النموذج مقابل حد التفاعل الثلاثي $X_1 X_2 X_3$ ولايقترح هذا الرسم (غير مبين هنـا) أي حاجة لحد تفاعل ثلاثي المتغيرات في نموذج الانحدار.

وبالإضافة إلى ذلك، فقد تم توفيق نموذج انحدار يتضمن X_1، X_2 و X_3 في حـدود من المرتبة الأولى وجميع حدود التفاعل ذات المتغيرين وحـد التفـاعل ثلاثي المتغـيرات. وقد أدت إضافة جميع حدود التفاعل إلى زيادة R^2 إلى 0.979 فقط بالمقارنـة مـع قيمـة 0.972 لِـR^2 في نموذج من المرتبـة الأولى في المتغـيرات المسـتقلة الثلاثـة. وبالاسـتناد إلى هذه النتائج وإلى نتائج سابقة تقرر أن لايتضمن نموذج الانحدار أي حدود تفاعل.

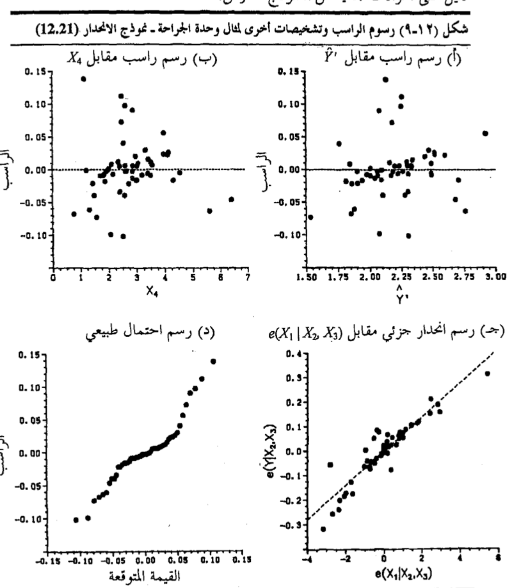
ويتضمن الشكل (١٢-٩) بعـض الرسـوم التشـخيصية الإضافيـة الـتي تم توليدهـا للتحقق من صلاحية نموذج المرتبة الأولى:

$$Y_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \qquad (12.21)$$

والنقاط التالية جديرة بالملاحظة:

١- لا يقدم رسم الراسب مقابل X_4 القيم التوفيقية، في الشكل (١٢-٩)أ، أي دليل على انحرافات جدية عن النموذج المفترض.

شكل (١٢-٩) رسوم الراسب وتشخيصات أخرى لمثال وحدة الجراحة - نموذج الانحدار (12.21)



٢- يبين الشكل (١٢-٩)ب رسم الرواسب مقابل X_4 المتغير المستقل غير الموجود في النموذج. ويبين هذا الرسم أنه لاحاجة لوجود وظيفة الكبد X_4 في النموذج المعد

للتنبؤ بالفترة التي يعيشها المريض بعد الجراحة.

٣- يبين رسم الانحدار الجزئي في الشكل (١٢-٩)جـ إمكانية وجود تأثير منحن للمتغير X_1. وقد تم تحرّي هذه الامكانية بإضافة حد تربيعي X_1^2 إلى النموذج وباستخدام التحويل اللوغاريتمي لـِ X_1. ولم يقدم أي من هذين التدبيرين العلاجيين تحسينا يُذكر.

٤- يبين رسم الاحتمال الطبيعي في الشكل (١٢-٩)د شيئا من الابتعاد عن الخطية. وكذلك فإن معامل الارتباط بين الرواسب المرتبة وبين قيمها المتوقعة تحت الطبيعية، وقيمته 0.959 أقل بقليل من القيمة الحرجة الموافقة لمستوي معنوية 0.01 في الجدول (٤-٣)، مما يشير إلى بعض الحيدان عن الطبيعية.

وعلى أي حال، فإن مشكلة عدم توافر التوزيع الطبيعي لم تعتبر هنا مشكلة جدية، إذ اقترح فحص الرواسب في الجدول (١٢-٤) ورسم الاحتمال الطبيعي في الشكل (١٢-٩)د أن الحيدان عن الطبيعية هو بصورة رئيسة في ذيلي التوزيع. وإلى حد ما، هناك من الرواسب القاصية أكثر مما هو متوقع تحت الطبيعية. ومع ذلك فهناك واحد فقط من الرواسب المعيّرة تقديرا، وعدّتها 54، يتجاوز في قيمته المطلقة 3، وجميع الرواسب القاصية الأخرى أصغر بكثير. وبالنظر إلى هذا المؤشر فإن الرواسب القاصية هي من النوع المعتدل، وطالما أن العينة من 54 مشاهدة هي عينة كبيرة الحجم نسبيا، فقد استتب الشعور بأن حيدانا معتدلا عن الطبيعية سوف لايكون له أي أثر جدي على ما سنستقرئه من نموذج الانحدار.

وقد دُرست الخطية المتعددة بحساب عوامل تضخم التباين:

المتغير	$(VIF)_k$
X_1	1.03
X_2	1.01
X_3	1.02

وكما يمكن رؤيته من هذه النتائج فإن الخطية المتعددة بين المتغيرات المستقلة الثلاثة ليست مشكلة.

ويتضمن الجدول (١٢-٤) نتائج بعض التحريات التشخيصية الخاصة بالمشاهدات القاصية والمشاهدات المؤثرة والقياسات المقدمة في الأعمدة من (١) إلى (٧) هي الرواسب e_i في (11.16)، وقيم العزم h_{ii} في (11.13)، والرواسب المعيرة تقديرا بصورة داخلية e_i^* في (11.19)، ورواسب الحذف d_i في (11.20)، ورواسب الحذف المعيرة تقديرا d_i^* في (11.23)، وقيم الـ (***DFFITS***) في (11.26)، وقياسات مسافة كوك D_i في (11.29). والنقاط التالية حول التشخيصات في الجدول (١٢-٤)، هي نقاط جديرة بالملاحظة:

١- المشاهدة 22 قاصية في Y، وذلك وفقا لكل من راسبها المعير تقديرا وراسب حذفها المعير تقديرا، فهي تنأى بأكثر من ثلاثة انحرافات معيارية. ويمكن اعتبار المشاهدة 27 على التخوم.

٢- وباستخدام $2p/n = 2(4)/54 = 1.148$ كدليل لتحديد المشاهدات القاصية في X، نلاحظ أن المشاهدات 13، 17، 28، 32 و38 هي مشاهدات قاصية وفقا لقيم عزومها. وبالمصادفة، فإن رسوم الجذع والورقة لمتغير واحد لاتحدد بوضوح المشاهدة 13 كمشاهدة قاصية. وما نشاهده هنا هو قيمة تحدد القاصية في حالة متغيرات متعددة.

٣- ولتحديد نفوذ أو تأثير المشاهدات 13، 17، 22، 27، 28، 32 و38 نعتبر قيم (***DFFITS***) ومسافة كوك لكل منها. ووفقا لكل من هذين المقياسين، فإن المشاهدة 38 هي المشاهدة الأكثر نفوذا من بينها بقيمة $(DFFITS)_{38} = 1.5282$ وقيمة $D_{38} = 0.5336$. وبالعودة إلى التوزيع F بـ 4 و 50 درجة من الحرية، نلاحظ أن قيمة كوك تقابل المئين 29. وهكذا يبدو أن نفوذ المشاهدة 38 ليس كبيرا إلى الحد الذي يستدعي تدبيرا علاجيا، وبالتالي فإن المشاهدات القاصية الأخرى لاتبدو بدورها مفرطة التأثير.

جدول (١٢-٤) تشخيصات مختلفة لمثال وحدة الجراحة - نموذج الانحدار (12.21)

(٧) D_i	(٦) $(DFFITS)_i$	(٥) d_i^*	(٤) d_i	(٣) e_i^*	(٢) h_{ii}	(١) e_i	رقم الحالة i
0.0001	0.0208	0.1262	0.0061	0.1274	0.0264	0.0059	1
0.0003	-0.0348	-0.1996	-0.0096	-0.2016	0.0295	-0.0093	2
0.0002	-0.0307	-0.1413	-0.0068	-0.1427	0.0452	-0.0065	3
0.0000	0.0112	0.0384	0.0019	0.0388	0.0789	0.0017	4
0.0048	-0.1376	-0.3668	-0.0185	-0.3700	0.1234	-0.0162	5
0.0035	-0.1178	-0.4572	-0.0223	-0.4609	0.0622	-0.0209	6
0.0000	-0.0127	-0.0572	-0.0028	-0.0578	0.0472	-0.0026	7
0.0105	-0.2047	-0.8617	-0.0416	-0.8639	0.0534	-0.0394	8
0.0393	-0.4131	-2.3041	-0.1053	-2.2108	0.0311	-0.1020	9
0.0109	-0.2080	-0.7466	-0.0365	-0.7499	0.0720	-0.0339	10
0.0043	0.1310	0.5708	0.0276	0.5747	0.0500	0.0263	11
0.0036	-0.1185	-0.3987	-0.0197	-0.4021	0.0812	-0.0181	12
0.0737	0.5469	1.3046	0.0658	1.2955	0.1495	0.0560	13
0.0036	0.1186	0.5181	0.0251	0.5219	0.0498	0.0238	14
0.0002	-0.0277	-0.1234	-0.0060	-0.1246	0.0478	-0.0057	15
0.0002	0.0244	0.1126	0.0055	0.1137	0.0448	0.0052	16
0.0062	-0.1558	-0.3709	-0.0190	-0.3741	0.1499	-0.0162	17
0.1004	-0.6454	-1.6917	-0.0833	-1.6610	0.1271	-0.0727	18
0.0002	0.0277	0.1433	0.0069	0.1447	0.0361	0.0067	19
0.0001	0.0201	0.1256	0.0060	0.1269	0.0250	0.0059	20
0.0006	-0.0492	-0.2854	-0.0137	-0.2880	0.0289	-0.0133	21
0.3641	1.3353	3.4951	0.1585	3.1588	0.1274	0.1383	22
0.0013	0.0714	0.1897	0.0096	0.1916	0.1240	0.0084	23
0.0007	0.0508	0.3317	0.0159	0.3347	0.0229	0.0155	24
0.0024	0.0975	0.4425	0.0214	0.4461	0.0463	0.0204	25
0.0003	0.0334	0.2365	0.0113	0.2387	0.0196	0.0111	26
0.0471	0.4571	2.5522	0.1153	2.4222	0.0311	0.1118	27
0.2209	-0.9546	-1.6027	-0.0861	-1.5781	0.2619	-0.0636	28
0.0028	-0.1042	-0.4667	-0.0226	-0.4704	0.0474	-0.0215	29
0.0503	0.4654	2.1945	0.1014	2.1153	0.0430	0.0970	30

جدول (١٢-٤) تتمة

0.0499	-0.4529	-1.5270	-0.0737	-1.5071	0.0808	-0.0677	31
0.0625	0.4997	0.9656	0.0510	0.9662	0.2113	0.0402	32
0.0002	-0.0305	-0.1894	-0.0091	-0.1912	0.0252	-0.0089	33
0.0037	0.1216	0.6425	0.0308	0.6463	0.0346	0.0298	34
0.0000	-0.0083	-0.0529	-0.0025	-0.0534	0.0239	-0.0025	35
0.0000	-0.0125	-0.0558	-0.0027	-0.0563	0.0479	-0.0026	36
0.1463	-0.7979	-2.3183	-0.1102	-2.2231	0.1059	-0.0985	37
0.5336	1.5282	2.3903	0.1271	2.2850	0.2902	0.0902	38
0.0164	0.2600	1.5873	0.0743	1.5637	0.0261	0.0723	39
0.0003	-0.0332	-0.1832	-0.0088	-0.1850	0.0317	-0.0085	40
0.0002	-0.0289	-0.1511	-0.0073	-0.1526	0.0353	-0.0070	41
0.0791	-0.5682	-1.4124	-0.0707	-1.3986	0.1393	-0.0608	42
0.0394	-0.3972	-1.0547	-0.0528	-1.0535	0.1242	-0.0462	43
0.0000	-0.0095	-0.0578	-0.0028	-0.0584	0.0265	-0.0027	44
0.0174	-0.2629	-0.8748	-0.0429	-0.8769	0.0828	-0.0394	45
0.0020	-0.0876	-0.3375	-0.0165	-0.3406	0.0631	-0.0155	46
0.0008	0.0574	0.2115	0.0104	0.2135	0.0686	0.0097	47
0.0059	0.1528	0.5040	0.0249	0.5078	0.0841	0.0228	48
0.0001	0.0197	0.1277	0.0061	0.1290	0.0233	0.0060	49
0.0004	-0.0378	-0.1173	-0.0058	-0.1185	0.0941	-0.0053	50
0.0000	-0.0053	-0.0192	-0.0009	-0.0193	0.0714	-0.0009	51
0.0000	-0.0057	-0.0166	-0.0008	-0.0168	0.1054	-0.0007	52
0.0005	0.0428	0.2603	0.0125	0.2627	0.0264	0.0122	53
0.0008	-0.0573	-0.1771	-0.0088	-0.1789	0.0948	-0.0080	54

وقد جرى أيضا فحص مباشر لتأثير المشاهدة 38 على الاستقراءات ذات الأهمية. وهنا فإن الاستقراءات ذات الأهمية الرئيسة هي في توفيق نموذج الانحدار لأن النموذج معد لاستخدامه في الحصول على تنبؤات ضمن مدى المشاهدات X. وبالتالي فقد قورنت كل قيمة توفيقية $\hat{Y}_i$ مستندة إلى المشاهدات الـ 54جميعها بالقيمة التوفيقية $\hat{Y}_{i(38)}$ عند حذف المشاهدة 38، عند توفيق نموذج الانحدار، ومتوسط مطلق الفروق النسبية المئوية:

$$\left|\frac{\hat{Y}_{i(38)} - \hat{Y}_i}{\hat{Y}_i}\right| 100$$

هو 0.2 بالمائة فقط، وأعلى مطلق فروق نسبية مئوية (وهو للمشاهدة 38) يبلغ 1.85 بالمائة. وهكذا لانجد للمشاهدة 38 تأثيرا على القيم التوفيقية متفاوتا إلى الحد الذي يدعو إلى عمل علاجي.

٤- والخلاصة أن التحليل التشخيصي قد حدد عددا من المشاكل المهمة ولكن أيا منها لم تُعتبر جدية إلى الحد الذي يستدعي القيام بعلاج.

وكنتيجة لذلك فقد اختير نموذج الانحدار (12.21) من بين النماذج المتنافسة المختلفة كي نقوم أخيرا بالتحقق من صحته. ونتائج توفيق هذا النموذج لبيانات الجدول (١٢-١) معطاة في الجدول (١٢-٥) في العمود الذي عنوانه "مجموعة بيانات بناء النموذج".

ودالة الانحدار المقدَّرة للنموذج الذي اختير هي:

$$\hat{Y}' = 0.84 + 0.0692X_1 + 0.00929X_2 + 0.00952X_3 \qquad (12.22)$$

جدول (١٢-٥) نتائج الانحدار مبنية على مجموعة بيانات بناء نموذج وعلى مجموعة بيانات التحقق من صحة نموذج وذلك من أجل مثال وحدة الجراحة - نموذج الانحدار (12.21)

الإحصاء	مجموعة بيانات بناء النموذج	مجموعة بيانات التحقق من صحة النموذج
b_0	.484	.501
$s\{b_0\}$	.043	.042
b_1	.0692	.0674
$s\{b_1\}$	.0041	.0050
b_2	.00929	.0101
$s\{b_2\}$	.00038	.00037
b_3	.00952	.00974
$s\{b_3\}$	.00031	.00030
SSE	.1099	.1056
$PRESS$	.1405	-
MSE	.0022	.0021
$MSPR$	-	.0072
R^2	.972	.978

(١٢-٦) التحقق من صحة نموذج

والخطوة النهائية في عملية بناء نموذج هي التحقق من صحة نموذج الانحدار الذي تم اختياره. وينطوي التحقق من صحة نموذج عادة على فحص النموذج بتطبيقه على بيانات مستقلة. وبما أن نماذج الانحدار تُستخدم، في الغالب، لأغراض متنوعة، فينبغي استخدام عدة طرق مختلفة للتحقق من صحة نموذج حيثما يكون ذلك ممكنا عمليا.

وفيما يلي ثلاث طرق أساسية للتحقق من صحة نموذج:

١- جميع بيانات جديدة لفحص النموذج وقدرته التنبؤية.

٢- مقارنة النتائج بتوقعات نظرية، وبنتائج تجريبية سابقة، وبنتائج محاكاة.

٣- الاستفادة من جز ء من العينة احتفظ بها لفحص النموذج وقدرته التنبؤية.

وسنناقش الآن كلا من هذه الطرق على التوالي.

تجميع بيانات جديدة لفحص نموذج

إن أفضل وسائل التحقق من صحة نموذج هي من خلال تجميع بيانات جديدة. والغرض من تجميع بيانات جديدة هو أن نتمكن من فحص ما إذا كان نموذج انحدار طورناه من بيانات سابقة لايزال قابلا للتطبيق على البيانات الجديدة. وإذا كان الأمر كذلك فسنطمئن إلى قابلية تطبيق النموذج فيما وراء البيانات التي بُني عليها.

وهناك تشكيلة من الطرق لفحص صحة نموذج انحدار بالاستناد إلى بيانات جديدة. إحداها هي إعادة تقدير صيغة النموذج الذي اختير سابقا مستخدمين البيانات الجديدة، ثم مقارنة معاملات النموذج الجديد وخواصه المختلفة من حيث اتساقها مع تلك الخاصة بنموذج الانحدار المبني على البيانات السابقة. وإذا كانت النتائج منسجمة (أو متسقة) فسيتوافر لنا دعم قوي للزعم بأن نموذج الانحدار الذي اختير قابل للتطبيق تحت ظروف أوسع من تلك المتصلة بالبيانات الأصلية.

وطريقة أخرى لفحص صحة نموذج هو استخدام البيانات الجديدة لإعادة تقدير جميع النماذج " الجيدة" التي اعتُبرت سابقا وذلك لرؤية ما إذا كان النموذج الذي اختير لايزال النموذج المفضل بالنسبة للبيانات الجديدة. وإذا كان الأمر كذلك فسنطمئن إلى فعالية النموذج الذي اختير تحت شروط جديدة.

وهناك طريقة ثالثة مصممة لمعايرة المقدرة التنبؤية لنموذج الانحدار الذي اختير إذ عندما نطوّر نموذج انحدار من بيانات معطاة، فمما لاشك فيه أن اختيار النموذج الذي اختير كان في معظمه بسبب توافقه توافقا جيدا مع البيانات المتوافرة. ومن أجل نتائج عشوائية مختلفة في مجموعة البيانات، كان يمكن الوصول إلى نموذج مختلف من حيث المتغيرات المستقلة التي يتضمنها و/أو مايتضمنه النموذج من صيغ دالية في هذه المتغيرات ومن حدود تفاعل بينها. وإحدى نتائج عملية تطوير النموذج هذه هي أن يميل متوسط مربعات الخطأ إلى أن لا يُفصح بحق عن خاصية التغير في التنبؤات المستقبلية MSE للنموذج المختار، وإنما ينحو إلى التخفيف منها.

وإحدى وسائل قياس القدرة التنبؤية لنموذج الانحدار المختار هي استخدام هذا النموذج للتنبؤ بكل مشاهدة في مجموعة البيانات الجديدة ثم حساب متوسط مربعات أخطاء التنبؤ، وسنرمز لها بـ MSPR (meansquare prediction error):

$$MSPR = \frac{\sum_{i=1}^{n^*}(Y_i - \hat{Y}_i)^2}{n^*} \qquad (12.23)$$

حيث:

Y_i قيمة متغير الاستجابة في المشاهدة من مجموعة البيانات الجديدة.

$\hat{Y}_i$ القيمة التنبؤية للمشاهدة i في مجموعة البيانات الجديدة محسوبة باستخدام النموذج الناتج عن البيانات الأصلية.

n^* عدد المشاهدات في مجموعة البيانات الجديدة.

وإذا كان متوسط مربعات خطأ التنبؤ **MSPR** قريب قربا معقولا من MSE الناتج عن توفيق نموذج الانحدار للبيانات الأصلية، فلا يكون MSE متوسط مربعات الخطأ لنموذج الانحدار المختار منحازا عندئذ انحيازا جديا ويعطي مؤشرا مناسبا للمقدرة التنبؤية للنموذج. وإذا كان متوسط مربعات خطأ التنبؤ أكبر بكثير من MSE، فينبغي الاعتماد على متوسط مربعات خطأ التنبؤ كمؤشر لقدرة نموذج الانحدار المختار على التنبؤ مستقبلا بصورة جيدة.

ملاحظة

عندما نجمع البيانات الجديدة تحت شروط تجريبية يمكن التحكم فيها، فمن المستحسن أن تتضمن هذه البيانات مشاهدات ذات أهمية رئيسة في مجال التحقق من المقدرة التنبؤية للنموذج. وإذا كان النموذج سيُستخدم للقيام بتنبؤات فوق كامل مدى المشاهدات X، فإحدى الإمكانيات هي أن تشمل البيانات الجديدة نقاطا تنتشر بانتظام فوق فضاء المتغيرات X.

المقارنة مع نتائج نظرية أو دلالة تجريبية أو نتائج محاكاة

في بعض الحالات يمكن أن تكون النظرية أو نتائج محاكاة، أو نتائج تجريبية سابقة، مفيدة في تحديد ما إذا كان النموذج المختار معقولا. وينبغي القيام بمقارنة معاملات انحدار وتنبؤات بتوقعات نظرية، بنتائج تجريبية سابقة، أو بنتائج محاكاة. ولسوء الحظ لايوجد، في الغالب، إلا القليل من النتائج النظرية التي يمكن استخدامها للتحقق من صحة نماذج انحدار.

تقسيم البيانات

الطريقة المفضلة إلى حد كبير للتحقق من صحة نموذج انحدار هي طريقة تجميع بيانات جديدة. إلا أن هذا لايكون، في الغالب، ممكنا ولا عمليا. وعندما تكون مجموعة البيانات كبيرة كبرا كافيا، فالبديل المعقول هو تقسيم البيانات إلى مجموعتين، الأولى، ويقال لها مجموعة بناء النموذج، تُستخدم لتطوير النموذج. ومجموعة البيانات الثانية، ويُقال لها مجموعة التنبؤ أو التقويم، تُستخدم لتقويم معقولية النموذج المختار ومقدرته التنبؤية. وتدعى هذه الطريقة، في الغالب التحقق المتصالب. ويشكل تقسيم البيانات، في واقع الأمر، محاكاة لإعادة الدراسة بصورة كاملة أو جزئية.

وتُستخدم مجموعة التقويم للتحقق من صحة النموذج بالطريقة نفسها التي تُستخدم فيها مجموعة بيانات جديدة. إذ يمكن إعادة تقدير معاملات الانحدار للنموذج المختار ومقارنتها مع المعاملات التي حصلنا عليها من مجموعة بناء النموذج من حيث اتساقها مع هذه المعاملات كما يمكن، أيضا، القيام بتنبؤات من أجل مجموعة التقويم

مستخدمين نموذج الانحدار الناشيء عن مجموعة بنـاء النمـوذج وذلـك لمعـايرة المقـدرة التنبؤية لنموذج الانحدار هذا من أجل بيانات جديدة. وعندما تكـون مجموعـة المعـايرة كبيرة بما يكفي فيمكن أيضا دراسة كيف تصير حال النماذج " الجيدة"، الـتي اعتـبرت في مرحلة اختيار النموذج، مع البيانات الجديدة.

وغالبا ماتُقسم مجموعة البيانات بالتساوي إلى مجموعتي بناء نموذج وتقويم. ومـن المهم، على أي حال، أن تكون مجموعة بناء النموذج كبيرة بمـا يكفـي لتطويـر نمـوذج يمكن الاعتماد عليه. ولنتذكر، في هذا السياق، أن عـدد المشـاهدات ينبغـي أن يكـون ستة إلى عشرة أضعاف عدد المتغيرات في مجموعـة المتغـيرات المسـتقلة. وهكـذا عندمـا يوجد في المجموعة عشرة متغـيرات مستقلة ينبغـي أن تتضمـن مجموعـة بنـاء النمـوذج مالايقل عن 60 إلى 100 مشاهدة. وإذا لم تكـن مجموعـة البيانـات بكاملهـا كبـيرة بمـا يسمح بتقسيمها مناصفة، فسيتوجب عندئذ جعل مجموعة التقويم أصغـر مـن مجموعـة بناء النموذج.

ويمكـن تقسـيم البيانـات بصـورة عشـوائية. والإمكانيـة الأخـرى هـي ملاءمـة المشاهدات في أزواج ووضع مشاهدة من كل زوج في إحدى مجموعتي التقسيم. وعند تجميع البيانات بصورة تتابعية مع الزمن، يكون من المفيد، غالبـا، اختيـار نقطـة زمنيـة تفصل بين مجموعتي البيانات. وبصورة عامة، نختار البيانات الـتي أُخـذت أولا لمجموعـة بنـاء النموذج والبيانات اللاحقـة لمجموعـة التقويـم. وعنـد وجـود تأثـيرات دوريـة أو موسمية في البيانات (مثلا، بيانات مبيعات) فينبغي أن تقع نقطة التقسيم حيث تتوازن الدورات.

ويمكن أن يخلق استخدام الزمن أو بعض الخواص الأخرى للبيانات لتقسـيمها إلى مجموعتين بعض المشاكل. فقد تختلف الشروط بالنسبة لمجموعتي البيانـات. وقـد تكـون مجموعة بيانات التقويم قد نتجت تحت ظروف سببية مختلفة عن تلك الخاصة بمجموعة بيانات النموذج، وعلى الوجه الآخـر يمكـن أن تمثـل مجموعـة بيانـات التقويـم امتـدادا لمجموعة بيانات النموذج (مثلا، بيانات مبيعات جمعت فـوق فـترة زمنيـة). وقـد تقـود شروط مفاضلة كهذه إلى نقص في صحة النموذج وتشير إلى الحاجة إلى توسيع نمـوذج الانحدار بحيث يصبح قابلا للتطبيق ضمن أفق أوسـع مـن الشـروط. وقـد يحتـاج المـرء

عندئذ إلى استخدام جزء من بيانات التقويم لتوسيع مـدى مجموعـة بيانـات النمـوذج، بينما لايزال يحتفظ إلى حد ما بمجموعة بيانات تقويم.

وأحد الطعون الممكنة لأسلوب تقسيم البيانات هو أن تباينات معاملات الانحدار المقدَّرة الناشئة عن مجموعة بيانات بناء النمـوذج سـتكون عـادة أكـبر ممـا لـو كنـا قـد حصلنا على المعاملات من توفيق مجموعة البيانات بكاملها. وعلى أي حال، فإن تلـك التباينات سوف لاتكـون بصـورة عامـة أكـبر بكثـير إذا كـانت مجموعـة بيانـات بنـاء النموذج كبيرة بصورة معقولة. ومهما يكن الأمر فالممارسة المعتـادة هـي أن نسـتخدم مجموعة البيانات بكاملها لتقدير نموذج الانحدار النهائي، وذلك حالما ننتهي من التحقق من صحة النموذج، مع أن الأكثر سلامة، من الناحية النظرية، هو الاحتفاظ بالنموذج التوفيقي الذي حصلنا عليه من مجموعة بيانات بناء النموذج بمفردها.

ملاحظة

من أجـل مجموعـة بيانـات صغـيرة حيـث يكـون تقسـيم البيانـات غـير عملـي، يمكـن استخدام طريقة الـ *PRESS* التي اعتبرناها سـابقا كمعيـار لاختيـار مجموعـات جزئيـة، كشكل من أشكال تقسيم البيانات بغية تقويـم دقـة تنبـؤات النمـوذج. ولنتذكـر أننـا نتنبأ، وفق هذه الطريقة بكل مشاهدة مستخدمين دالة انحـدار المربعـات الدنيـا الناشـئة عن الـ $n - 1$ مشاهدة الباقية. وسيقترح اقـتراب قيمـتي *PRESS* و*SSE* مـن بعضهمـا امكانية اعتبار *MSE* كمؤشر مشروع إلى حد ما، للمقدرة التنبؤية للنموذج المختار.

مثال

في مثال وحدة الجراحة اختير نموذج الانحدار التوفيقـي (12.22) أخـيرا كنمـوذج يُـراد التحقق من صحته. ويبين الجـدول (١٢ـ٥) معـاملات الانحـدار المقـدَّرة، وانحرافاتهـا المعيارية المقدَّرة، وبعض الإحصاءات الأخرى المتعلقة بالنموذج التوفيقي.

وقد ذكرنا سابقا بعض الدلائل على صحة هـذا النمـوذج التوفيقـي وفـق معايـير داخلية. ولاحظنا في الجدول (١٢-٣) قرب. $PRESS = 0.1405$ و$SSE = 0.1099$ مـن أجل هذا النموذج. وقيمة *PRESS* ستكون دائما أكبر من *SSE* لأن توفيق الانحدار مع

الغاء مشاهدة لا يمكن أن يكون في جودة انحدار توفيقي يتناول جميع المشاهدات بما فيه المشاهدة i. وقيمة لِـ $PRESS$ قريبة من SSE كما هو الحال هنا، تدعم صحة نموذج الانحدار التوفيقي كما تدعم اتخاذ MSE كمؤشر للمقدرة التنبؤية لهذا النموذج.

وللتحقق من صحة نموذج الانحدار المختار بمعايير خارجية، فقد تمّ الاحتفاظ بأربع وخمسين مشاهدة كمجموعة بيانات تقويم. وبيانات هذه المشاهدات معروضة في الجدول (١٢-٦). ومصفوفة الارتباط لهذه البيانات (غير مبينة هنا) مشابهة تماما لتلك المبينة في الجدول (١٢-٢) لمجموعة بيانات بناء النموذج. ويبين الجدول (١٢-٥) معاملات الانحدار المقدَّرة، وانحرافاتها المعيارية المقدَّرة، وبعض الإحصاءات الأخرى، وذلك عند توفيق نموذج الانحدار (12.21) لمجموعة بيانات التقويم. لاحظ التوافق الجيد بين المجموعتين من معاملات الانحدار المقدَّرة، وبين قيمتي MSE في الحالتين وبين قيمتي R^2 في الحالتين.

جدول (١٢-٦) مجموعة بيانات تقويم ـ مثال وحدة الجراحة.

رقم المشاهدة i	درجة تخثر الدم X_{i1}	دليل التناذر X_{i2}	اختبار وظيفة الانزيم X_{i3}	اختبار وظيفة الكبد X_{i4}	Y_i
1	7.1	23	78	1.93	2.0326
2	4.9	66	91	3.05	2.4086
3	6.4	90	35	1.06	2.2177
4	5.7	35	70	2.13	1.9078
5	6.1	42	69	2.25	2.0035
6	8.0	27	83	2.03	2.0945
7	6.8	34	51	1.27	1.7652
8	4.7	63	36	1.71	1.7925
9	7.0	47	67	1.60	2.1292
10	6.7	69	65	2.91	2.2295
11	6.7	46	78	3.26	2.1524
12	5.8	60	86	3.11	2.3188
13	6.7	56	32	1.53	1.9039
14	6.8	51	58	2.18	2.0508
15	7.2	95	82	4.68	2.6525
16	7.4	52	67	3.28	2.2053
17	5.3	53	62	2.42	1.9246
18	3.5	58	84	1.74	2.1541

جدول (١٢-٦) تتمة

رقم المشاهدة i	درجة تخثر الدم X_{i1}	دليل التناذر X_{i2}	اختبار وظيفة الانزيم X_{i3}	اختبار وظيفة الكبد X_{i4}	Y_i
19	6.8	74	79	2.25	2.4970
20	4.4	47	49	2.42	1.7237
21	7.0	66	118	4.69	2.8339
22	6.7	61	57	3.87	2.1282
23	5.6	75	103	3.11	2.6884
24	6.9	58	88	3.46	2.4284
25	6.2	62	57	1.25	2.0261
26	4.7	97	27	1.77	2.0843
27	6.8	69	60	2.90	2.2826
28	6.0	73	58	1.22	2.2073
29	5.9	50	62	3.19	2.0443
30	5.5	88	74	3.21	2.4863
31	3.8	55	52	1.41	1.9037
32	4.3	99	83	3.93	2.6647
33	6.6	48	54	2.94	1.9071
34	6.2	42	63	1.85	1.9093
35	5.0	60	105	3.17	2.4389
36	5.8	62	82	3.18	2.3343
37	4.7	42	10	0.28	1.3379
38	5.7	70	59	2.28	2.1996
39	4.7	64	48	1.30	1.8795
40	7.8	74	40	2.58	2.1504
41	2.9	43	32	0.94	1.4330
42	4.9	72	90	3.51	2.4381
43	4.6	73	57	2.82	2.1075
44	5.9	78	70	4.28	2.2843
45	4.6	69	70	3.17	2.1615
46	6.1	53	52	1.84	2.0558
47	5.9	88	98	3.33	2.7249
48	4.7	66	68	1.80	2.0520
49	10.4	62	85	4.65	2.6810
50	5.8	70	64	2.52	2.2604
51	5.4	64	81	1.36	2.2553
52	6.9	90	33	2.78	2.1745
53	7.9	45	55	2.46	2.0224
54	4.5	68	60	2.07	2.1413

ولمعايرة المقدرة التنبؤية لنموذج الانحدار التوفيقي من مجموعة بناء النموذج، فقد تمَّ حساب متوسط مربعات خطأ التنبؤ *MSPR* في (12.23) للمشاهدات الواردة في مجموعة بيانات التقويم في الجدول (١٢-٦) فكانت $MSPR = 0.0072$. وبصورة عامة سيكون متوسط مربعات خطأ التنبؤ أكبر من *MSE* الخاصة بمجموعة بيانات بناء النموذج لأن المشاهدات التي انطوت عليها مجموعة بيانات التقويم كانت مشاهدات جديدة كليا. وتتضمن حقيقة عدم اختلاف *MSPR* هنا اختلافا كبيرا عن *MSE*، أن متوسط مربعات الخطأ *MSE* المستند إلى مجموعة بيانات بناء النموذج هو مؤشر مشروع بصورة معقولة للمقدرة التنبؤية لنموذج الانحدار التوفيقي.

وتدعم نتائج التحقق من صحة النموذج هذه ملاءمة النموذج المختار.

تعليقات

١- تتوافر خوارزميات لتقسيم البيانات بحيث يكون لمجموعتي البيانات خواص إحصائية متشابهة. ونشير القارئ إلى المرجع [12.6] حيث يجد مناقشة لهذه المسألة ولمسائل أخرى تتعلق بالتحقق من صحة نموذج.

٢- اقتُرحت تحسينات لمسألة تقسيم البيانات. وعلى سبيل المثال، نقوم في طريقة التحقق المتصالبة المضاعفة، ببناء النموذج من أجل كل من نصفي البيانات المقسمة، ونختبره مستخدمين النصف الآخر من البيانات. وهكذا نحصل على مقياسين للاتساق وللمقدرة التنبؤية من النموذجين اللذين قمنا بتوفيقهما.

٣- عندما تكون مجموعة البيانات صغيرة. اقتُرحت أساليب *PRESS* متنوعة، يُحتفظ فيها بـ m مشاهدة للتحقق، وتُستخدم الـ $n - m$ مشاهدة الباقية لبناء النموذج. ويناقش المرجع [12.7] هذه الأساليب، بالإضافة إلى مسائل تعالج التقسيم الأمثل لمجموعة بيانات.

٤- عندما لاتتنبأ نماذج انحدار، مبنية على بيانات مشاهد، بصورة جيدة خارج مجال المشاهدات X في مجموعة البيانات، يكون السبب المعتاد هو وجود خطية متعددة بين المتغيرات المستقلة. وكما ذكرنا في الفصل الحادي عشر فمن بين الحلول الممكنة

لهذه الصعوبة نجد تحليل الحافة أو طرق تقدير منحازة أخرى.

(١٢-٧) ملاحظات ختامية

إن بناء نموذج انحدار فعّال ومناسب هو مهمة معقدة. وتقع مفاتيح النجاح في الصياغة المناسبة للمسألة، وفي تجميع قدر كاف من البيانات ذات النوعية العالية، واختيار متغيرات مهمة وصيغة مناسبة للنموذج.

مراجع ورد ذكرها

[12.1] Daniel, C. and Wood, F. *Fitting Equations to Data,* 2nd ed. New York : Wiley-Interscience, 1980.

[12.2] Dixon, W. J. (chief editor). BMDP *Statistical Software Manual*, vols. 1 and 2. Berkeley, Calif. : University of Califomai Press, 1988.

[12.3] Freedman, D. A. "A Note on Screening Regression Equations." *The American Statistician* 37 (1983), 152-55.

[12.4] Pope, P. T. and Webster, J. T. "The Use of an F-Statistic in Stepwise Regression." *Technometrics* 14 (1972), 327 - 40.

[12.5] Manted, N. "Why Stepdown Procedures in Variable Selection." *Technometrics* 12 (1970), 621 - 25.

[12.6] Snee, R. D. "Validation of Regresion Models : Methods and examples." *Technometrics* 19 (1977), 415 - 28.

[12.7] Stone, M."Cross-validatory Choice and Assessment of Statistical Prediction." *Journal of the Royal Statistical Society* B 36 (1974), 111 - 47.

مسائل

(١٢-١) صرح متحدث: "في التجارب المصممة بشكل جيد والمتضمنة لمتغيرات مستقلة كمية، لاضرورة لطريقة نخفض فيها عدد المتغيرات المستقلة بعد الحصول على البيانات". ناقش.

(١٢-٢) يرغب عميد مدرسة دراسات عليا في التنبؤ بالمعدل التراكمي في مرحلة الدراسات العليا لمتقدمين جدد. عدّد دزينة من المتغيرات التي يمكن أن تكون متغيرات مستقلة مفيدة.

(١٢-٣) درس باحثان العوامل المؤثرة في عدد المصطافين في شواطيء تابعة للقطاع الخاص على بحيرة أونتاريو، وجمعا معلومات حول المصطافين و 11 متغيرا تفسيريا لاثنين وأربعين من الشواطيء. وقد دُرس موسما اصطياف بطقس حار

نسبيا وبارد نسبيا، على الترتيب، ويُراد الآن استخدام خوارزمية" أفضل" مجموعات جزئية لتخفيض عدد المتغيرات المستقلة في نموذج الانحدار النهائي.

أ - هل ينبغي القيام بتخفيض المتغيرات للموسمين معا أم تخفيضها لكل موسم على حدة؟ اشرح المشاكل التي تنطوي عليها المسألة وكيف يمكنك معالجتها.

ب - هل ستختار طريقة اختيار المجموعات الجزئية " الأفضل" تلك المتغيرات المستقلة الأكثر أهمية، بالمعنى السببي للكلمة، في تحديد عدد المصطافين؟

(١٢-٤) في انحدار الخطوة فخطوة إلى الأمام ما هي فائدة استخدام قيمة عالية نسبيا لـ "F - للدخول" من أجل إضافة متغيرات؟ ما هي فائدة استخدام قيمة صغيرة لـ"F - للدخول"؟.

(١٢-٥) في انحدار الخطوة فخطوة إلى الأمام، لماذا يجب أن لاتتجاوز قيمة "F - للحذف" الخاصة بحذف متغيرات أبدا قيمة "F - للدخول" الخاصة بإضافة متغيرات؟

(١٢-٦) ارسم مخطط تدفق لكل من طرق الاختيار التالية: (١) انحدار الخطوة فخطوة إلى الأمام، (٢) الاختيار بالإضافة و(٣) الاختيار بالحذف.

(١٢-٧) صرح مهندس: ينبغي القيام بتخفيض المتغيرات المستقلة دائما باستخدام أسلوب الانحدار خطوة فخطوة إلى الأمام الموضوعي" ناقش.

(١٢-٨) صرح أحد الملتحقين بمقرر قصير في نمذجة الانحدار: "نادرا ما أرى التحقق من صحة نموذج انحدار مذكورا في أبحاث منشورة، ولذلك فلابد أنه، في حقيقة الأمر، لا يشكل مركبة مهمة في مسألة بناء نموذج". علّق.

(١٢-٩) بالإشارة إلى مسألة **راحة المريض** (٧-١٧). يرغب مدير المستشفى بتحديد أفضل مجموعة جزئية من المتغيرات المستقلة للتنبؤ براحة المريض.

أ - اذكر المجموعة الجزئية من المتغيرات المستقلة التي يمكن أن توصي بها كأفضل مجموعة جزئية للتنبؤ براحة المريض، وذلك وفقا لكل

معيـــار مـــن المعايـــير التاليـــة: (١) R_p^2 (٢) MSE_p (٣) C_p (٤) $PRESS_p$ ادعم توصياتك برسوم بيانية مناسبة.

ب ـ هل تحدد المعايير الأربعة في الجزء (أ)، المجموعة الجزئية الأفضل نفســها؟ هل يحدث هذا دائما؟

جـ ـ هل لانحدار الخطوة فخطوة إلى الأمام، كطريقـة غربلـة، أيـة مزايـا هنـا فوق طريقة جميع الانحدارات الممكنة؟

(١٢ـ١٠) **كسوة السقف بالألواح**. فيمـا يلـي بيانـات عـن مبيعـات السـنة الماضيـة (بآلاف المربعات) في ست وعشـرين منطقـة بيـع وذلـك لمنتـج ألـواح أسـفلتية لكسوة السقوف. وتبيان أيضا للنفقـات التشـجيعية (X_1 بـآلاف الـدولارات)، عدد الحسابات المصرفية العاملة X_2، عدد الأصناف المنافسة X_3، وإمكانات المنطقة (X_4، مرمزة) وذلك لكل منطقة من المناطق.

أ ـ قم بإعداد رسوم نقطية لكل من المتغيرات المستقلة. هل هناك أية نواح تستحق الذكر على هذه الرسوم؟ علّق.

ب ـ قم بإعداد رسوم انتشار منفصلة للمبيعات Y مقابل كل من المتغـيرات المستقلة الأربعة. هل تقترح رسوم الانتشار علاقات خطية أو منحنية بين المتغير التابع وكل من المتغيرات المستقلة ؟ ناقش.

جـ ـ أوجد مصفوفة ارتباط المتغيرات X. هل تتضح من هذه المصفوفـة أيـة مشاكل خطّية متعددة جدية؟. اشرح.

د ـ قم بتوفيق دالة انحدار متعدد تتضمن المتغيرات المستقلة الأربعة جميعها بحدود من المرتبة الأولى.

هـ ـ أوجد عوامـل تضخـم التبـاين للنمـوذج التوفيقـي في (د). هـل هنـاك مايشير إلى مشاكل خطية متعددة جدية هنا ؟ اشرح.

و ـ أوجد الرواسم وارسمها بصورة منفصلة في مقابل $\hat{Y}$، وكل من المتغيرات المستقلة. وقم أيضا بإعداد رسم احتمال طبيعي للرواسب. وعلى أساس من هذه الرسومات. هل ينبغي القيام بأية تعديلات في نموذج الانحدار؟

المنطقة i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	Y_i
1	5.5	31	10	8	79.3
2	2.5	55	8	6	200.1
3	8.0	67	12	9	163.2
4	3.0	50	7	16	200.1
5	3.0	38	8	15	146.0
6	2.9	71	12	17	177.7
7	8.0	30	12	8	30.9
8	9.0	56	5	10	291.9
9	4.0	42	8	4	160.0
10	6.5	73	5	16	339.4
11	5.5	60	11	7	159.6
12	5.0	44	12	12	86.3
13	6.0	50	6	6	237.5
14	5.0	39	10	4	107.2
15	3.5	55	10	4	155.0
16	8.0	70	6	14	291.4
17	6.0	40	11	6	100.2
18	4.0	50	11	8	135.8
19	7.5	62	9	13	223.3
20	7.0	59	9	11	195.0
21	6.7	53	13	5	73.4
22	6.1	38	13	10	47.7
23	3.6	43	9	17	140.7
24	4.2	26	8	3	93.5
25	4.5	75	8	19	259.0
26	5.6	71	4	9	331.2

(١٢-١١) بالعودة إلى مسألة **كسوة السقف بالألواح** (١٢-١٠).

أ ـ مستخدما فقط حدودا من المرتبة الأولى في المتغيرات المستقلة، أوجد نماذج انحدار المجموعات الجزئية الثلاث الأفضل وفقا للمعيار C_p.

ب ـ هل هناك انحياز قليل نسبيا في كل من هذه النماذج الثلاثة.

(١٢-١٢) بالعودة إلى مسألتي **كسوة السقف بالألواح** (١٢-١٠)، (١٢-١١). نريد تقويم النموذج الجزئي من المرتبة الأولى المتضمن X_2 و X_3 تقويما تفصيليا.

أ - أوجد الرواسب وارسمها بصورة منفصلة مقابل $\hat{Y}$ وكل من المتغيرات المستقلة الأربعة، ومقابل الحد الجدائي $X_2 X_3$. وعلى أساس من هذه الرسومات، هل ينبغي استقصاء أية تعديلات في نموذج الانحدار؟

ب - قم بإعداد رسومات انحدار جزئي منفصلة مقابل $e(X_2|X_3)$ و $e(X_3|X_2)$. هل تقترح هذه الرسومات مسوّغا لأية تعديلات في صيغة النموذج؟

جـ - قم بإعداد رسم احتمال طبيعي للرواسب. وأوجد أيضا معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية. إختبر معقولية افتراض الطبيعية مستخدما الجدول (٤-٣) و $\alpha = 0.05$. ماذا تستنتج؟

د - أوجد العناصر القطرية لمصفوفة القبعة. ومستخدما قاعدة الإبهام $2p/n$ حدد أية مشاهدات قاصية في X. هل يتفق ماوجدته هنا مع ماوجدته في المسألة (١٢-١٠)أ؟ هل ينبغي أن يكون هناك اتفاق؟ ناقش.

هـ - أوجد رواسب الحذف المعيّرة تقديرا وحدد أية مشاهدات قاصية في Y.

و - تبدو المشاهدة 8 قاصية في Y. أوجد قيم $DFFITS$، $DFBETAS$، ومسافة كوك لهذه المشاهدة لتثمين نفوذها. ماذا تستنتج؟

(١٢-١٣) **البراعة في عمل.** قام مسؤول في دائرة شؤون الموظفين في وكالة حكومية بالإشراف على أربعة اختبارات أهلية مبتكرة حديثا لكل من 25 متقدما لوظائف مكتبية بسيطة في الوكالة. ولأغراض الدراسة، تم قبول جميع المتقدمين الخمسة وعشرين في وظائف وذلك بصرف النظر عن درجاتهم في الاختبار. وبعد فترة مراقبة، تم ترتيب كل متقدم من حيث براعته في العمل وكانت الدرجات في الاختبارات الأربعة (X_1، X_2، X_3 وX_4) ودرجة البراعة في العمل Y للمتقدمين الخمسة وعشرين كما يلي:

الشخص	درجة الاختبار				درجة البراعة في العمل
i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	Y_i
1	86	110	100	87	88
2	62	97	99	100	80
3	110	107	103	103	96
4	101	117	93	95	76
5	100	101	95	88	80
6	78	85	95	84	73
7	120	77	80	74	58
8	105	122	116	102	116
9	112	119	106	105	104
10	120	89	105	97	99
11	87	81	90	88	64
12	133	120	113	108	126
13	140	121	96	89	94
14	84	113	98	78	71
15	106	102	109	109	111
16	109	129	102	108	109
17	104	83	100	102	100
18	150	118	107	110	127
19	98	125	108	95	99
20	120	94	95	90	82
21	74	121	91	85	67
22	96	114	114	103	109
23	104	73	93	80	78
24	94	121	115	104	115
25	91	129	97	83	83

أ ـ قم بإعداد رسومات جذع وورقة منفصلة لكل مجموعة من المجموعات الأربع لدرجات اختبار أهلية مبتكرحديثا، هل هناك نواح تستحق الذكر في هذه الرسومات ؟ علق.

ب ـ قم بإعداد رسوم انتشار منفصلة لدرجات البراعة Y مقابل كل من المتغيرات المستقلة. ماذا تقترح رسوم الانتشار حول طبيعة العلاقة الدالية بين المتغير التابع Y وكل من المتغيرات المستقلة؟

جـ ـ أوجد مصفوفة الارتباط للمتغيرات X هل تتضح من هذه المصفوفة أية مشاكل خطية متعددة جدية ؟ اشرح.

د ـ قم بتوفيق دالة انحدار متعدد تتضمن المتغيرات المستقلة الأربعة في حدود من المرتبة الأولى.

هـ ـ أوجد عوامل تضخم التباين لنموذج الانحدار التوفيقـي في الجـزء (د).
هل هناك مؤشرات لوجود مشاكل خطية متعددة جدية ؟ إشرح.

و ـ أوجد الرواسب وارسمها بصورة منفصلة مقابل $\hat{Y}$ وكل من المتغيرات المستقلة. أوجد أيضا رسم احتمال طبيعـي للرواسـب. وعلـى أسـاس من هذه الرسوم هل ينبغي القيام بأية تعديلات في نموذج الانحدار؟

(١٢ـ١٤) بالعودة إلى مسألة **البراعة في عمل** (١٢ـ١٣).

أ ـ مستخدما فقط حدود المرتبة الأولى للمتغيرات المستقلة في جملـة المتغـيرات X المرشحة، أوجد نماذج انحدار المجموعـات الجزئيـة الأربعـة الأفضـل وفقا للمعيار R^2 المعدل.

ب ـ وبما أنـه توجـد فـروق بسيطة نسـبيا في R_a^2 بـين نمـاذج المجموعـات الجزئية الأربع الأفضل. ماهو المعيار الآخـر الـذي يمكنـك اسـتخدامه للمساعدة في اختيار النموذج الأفضل؟ ناقش.

(١٢ـ١٥) بالعودة إلى مسألتي **البراعة في عمل** (١٢ـ١٣) و(١٢ـ١٤). نريد تقويما تفصيليا لنموذج المجموعة الجزئية المتضمن لحدود من المرتبة الأولى فقط في X_1 وX_3.

أ ـ أوجد الرواسب وارسمها بصورة منفصلة مقابل $\hat{Y}$ وكل من المتغيرات المستقلة الأربعـة، والحـد الجدائي $X_1 X_3$ وعلـى أسـاس مـن هـذه الرسومات هل ينبغي استقصاء أية تعديلات في نموذج الانحدار؟

ب ـ قم بإعداد رسوم انحدار جزئي منفصلة مقابل $e(X_1 \mid X_3)$ و$e(X_3 \mid X_1)$.
هل تقترح هذه الرسومات مسوغا لأية تعديلات في صيغة النموذج؟

جـ ـ قم بإعداد رسم احتمال طبيعي للرواسب. وأوجد أيضـا معـامل الارتبـاط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية. اختبر معقولية افـتراض الطبيعية مستخدما الجدول (٣ـ٤) و$\alpha = 0.01$. ماذا تستنتج؟

د ـ أوجد العناصر القطرية في مصفوفة القبعة. مستخدما قاعدة الإبهام $2p / n$ حدّد أية مشاهدات قاصية في X. هل يتفق ماوجدته مع ماوجدت في المسألة (١٢-١٣)أ؟ هل ينبغي أن يكون هناك اتفاق؟ علّق.

هـ ـ أوجد رواسب الحذف المعيّرة وحدّد أية مشاهدات قاصية في Y.

و ـ تبدو المشاهدتان 7 و 18 قاصيتين بصورة معتدلة بالنسبة لقيم X، كما تبدو المشاهدة 16 قاصية بالنسبة لقيم Y. أوجد قيم *DFFITS*، *DFBETAS*، ومسافة كوك لهذه المشاهدات وذلك لتثمين نفوذها. ماذا تستنتج؟

(١٢-١٦) **ضغط الرئة.** كثيرا ما تقود زيادة ضغط الدم الشرياني في الرئتين إلى تطور حالة قصور قلب في مرضى يعانون من مرض رئوي سادّ مزمن (*COPD*). والطريقة القياسية لتحديد ضغط الرئة الشرياني هي طريقة باضعة، صعبة تقنيا، وتنطوي على بعض الخطورة للمريض. وطريقة التصوير الإشعاعي النووي هي طريقة غير باضعة وأقل خطورة لتقدير الضغط الشرياني في الرئتين. ولدراسة القدرة التنبؤية لهذه الطريقة، جمع اختصاصي أشعة بيانات عن 19 حالة خفيفة إلى معتدلة من مرضى الـ *COPD*. وتتضمن البيانات التالية القياس الباضع للضغط الشرياني الرئوي الانقباضي Y وثلاثة متغيرات تنبؤ مرشحة وهي متغيرات غير باضعة. وقد تم الحصول على اثنين منها باستخدام طريقة التصوير الشعاعي ـ معدل إفراغ الدم إلى غرفة الضخ في القلب X_1، ومعدل دفع الدم الذي يُضخ من القلب إلى الرئتين X_2. ويقيس المتغير المستقل الثالث غاز الدم X_3.

أ ـ قم بإعداد رسوم نقطية منفصلة لكل من المتغيرات المستقلة الثلاثة. هل هناك أية نواح تستحق الملاحظة في هذه الرسوم؟ علّق.

ب ـ قم بـإعداد رسـوم انتشـار منفصلـة لـ Y مقـابل كـل مـن المتغـيرات المستقلة الثلاثة ماذا تقترح رسوم الانتشار هذه حـول طبيعـة العلاقـة الدالية بين Y وكل من المتغيرات المستقلة؟

الشخص i	X_{i1}	X_{i2}	X_{i3}	Y_i	الشخص i	X_{i1}	X_{i2}	X_{i3}	Y_i
1	45	36	45	49	11	37	37	55	31
2	30	28	40	55	12	29	34	47	49
3	11	16	42	85	13	26	32	28	38
4	30	46	40	32	14	38	45	30	41
5	39	76	43	26	15	38	99	26	12
6	42	78	27	28	16	25	38	47	44
7	17	24	36	95	17	27	51	44	29
8	63	80	42	26	18	37	32	54	40
9	25	12	52	74	19	34	40	36	31
10	32	27	35	37					

جـ ـ أوجد مصفوفة الارتباط للمتغيرات X، هل تتضح، من هذه المصفوفة، أية مشاكل خطية متعددة جدية؟. اشرح.

د ـ قم بتوفيق دالة انحدار متعدد تتضمن المتغيرات المستقلة الثلاثة في حدود من المرتبة الأولى.

هـ ـ أوجد عوامل تضخم التباين لنموذج الانحدار التوفيقـي في الجـزء (د). هـل هنـاك مايشـير إلى وجـود مشـاكل خطيـة متعـددة جديـة هنـا؟ اشرح.

و ـ أوجد الرواسب وارسمها بصورة منفصلـة مقـابل $\hat{Y}$ ومقـابل كـل مـن المتغيرات المستقلة الثلاثة، ومقابل كل حـد تفـاعل بـين عـاملين. قـم أيضـا بـإعداد رسـم احتمـال طبيعي للرواسـب. واسـتنادا إلى هــذه الرسوم، هل ينبغي تعديل النموذج؟ ناقش.

(١٢-١٧) بالإشارة إلى مسألة **ضغط الرئة** (١٢-١٦).

أ ـ مستخدما حـدودا مـن المرتبـة الأولى ومـن المرتبـة الثانيـة لكـل مـن المتغيرات المستقلة الثلاثة (معـبرا عنهـا كانحرافـات عـن المتوسـط) في

جملة المتغيرات X المرشحة (بما في ذلك جداءات الحدود من المرتبة الأولى)، أوجد وفقا للمعيار R_a^2 نماذج انحدار أفضل ثلاث مجموعات جزئية متسلسلة ويجب أن تتضمن المجموعات الجزئية المتسلسلة الحد من المرتبة الأولى لمتغير مستقل (مثلا، x_1) إذا تضمن النموذج حدا من المرتبة الثانية يتضمن ذلك المتغير (مثلا، x_1^2 أو x_1x_2).

ب ـ هل يوجد فارق كبير في R^2 بين النماذج الثلاثة الأفضل؟

(١٢-١٨) بالإشارة إلى مسألتي **ضغط الرئة** (١٢-١٦) و(١٢-١٧)، نريد تقويما تفصيليا لنموذج الانحدار المتضمن لحدود من المرتبة الأولى في X_1 و X_2 والحد الجدائي X_1X_2.

أ ـ أوجد الرواسب وارسمها بصورة منفصلة مقابل $\hat{Y}$ وكل من المتغيرات المستقلة الثلاثة وعلى أساس من هذه الرسوم، هل تنبغي محاولة القيام بأية تعديلات إضافية في النموذج؟

ب ـ قم بإعداد رسم احتمال طبيعي للرواسب. أوجد أيضا معامل ارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية. هل يبدو افتراض الطبيعية معقولا هنا؟

جـ ـ أوجد عوامل تضخم التباين. هل هناك أية مؤشرات لوجود مشاكل جدية لخطية متعددة ؟ اشرح.

د ـ أوجد العناصر القطرية لمصفوفة القبعة. ومستخدما قاعدة الإبهام $2p / n$ حدد أية مشاهدات قاصية في X. هل ينسجم ماوجدته هنا مع ماوجدته في المسألة (١٢-٦)أ؟ هل ينبغي وجود انسجام؟ ناقش.

هـ ـ أوجد رواسب الحذف المعيرة تقديرا وحدد أية مشاهدات قاصية في Y.

و ـ المشاهدات 3، 8 و15 هي مشاهدات قاصية باعتدال في قيم X، والمشاهدة 7 قاصية نسبيا في قيمة Y. أوجد قيم *DFFITS*، *DFBETAS*، ومسافة كوك لهذه المشاهدات بغية تثمين نفوذها. ماذا تستنتج؟

(١٢-١٩) **وظيفة الكُلية**. تصفية الكرياتينين Y هو قياس مهم لوظيفة الكلية ولكن من الصعب الحصول عليه في ترتيبات العيادات لأنه يتطلب تجميع البول لفترة 24 ساعة. ولتحديد ما إذا كان يمكن التنبؤ بهذا القياس من بعض البيانات المتوافرة بسهولة، قام اختصاصي كُلية بجمع البيانات التالية من 33 ذكرا. والمتغيرات المستقلة هي تركيز الكرياتينين في الدم X_1، العمر X_2، والوزن X_3.

Y_i	X_{i3}	X_{i2}	X_{i1}	الشخص i
132	71	38	.71	1
53	69	78	1.48	2
50	85	69	2.21	3
82	100	70	1.43	4
110	59	45	.68	5
100	73	65	.76	6
68	63	76	1.12	7
92	81	61	.92	8
60	74	68	1.55	9
94	87	64	.94	10
105	79	66	1.00	11
98	93	49	1.07	12
112	60	43	.70	13
125	70	42	.71	14
108	83	66	1.00	15
30	70	78	2.52	16
111	73	35	1.13	17
130	85	34	1.12	18
94	68	35	1.38	19
130	65	16	1.12	20
59	53	54	.97	21
38	50	73	1.61	22
65	74	66	1.58	23
85	67	31	1.40	24
140	80	32	.68	25
80	67	21	1.20	26
43	72	73	2.10	27
75	67	78	1.36	28
41	60	58	1.50	29
120	107	62	.82	30
52	75	70	1.53	31
73	62	63	1.58	32
57	52	68	1.37	33

أ - قم بإعداد رسوم نقطية منفصلة لكل من المتغيرات المستقلة الثلاثة، هل هناك أية نواح تستحق الملاحظة في هذه الرسوم؟ علّق.

ب - قم بإعداد رسوم انتشار منفصلة لـ Y مقابل كل من المتغيرات المستقلة الثلاثة. ماذا تقترح هذه الرسومات حول طبيعة العلاقة الداليّة بين المتغير التابع وكل من المتغيرات المستقلة ؟ ناقش.

جـ ـ أوجد مصفوفة الارتباط للمتغيرات X هل تتضح، من هذه المصفوفة، أية مشاكل جدية لخطية متعددة؟ اشرح.

د ـ قم بتوفيق دالة الانحدار المتعدد المتضمنة للمتغيرات المستقلة الثلاثة في حدود من المرتبة الأولى.

هـ ـ أوجد عوامل تضخم التباين لنموذج الانحدار الذي قمت بتوفيقه في الجزء (د). هل هناك مؤشرات لوجود مشاكل جدية لخطية متعددة هنا؟ اشرح.

و ـ أوجد الرواسب وارسمها بصورة منفصلة مقابل $\hat{Y}$ وكل من المتغيرات المستقلة. قم أيضا بإعداد رسم احتمال طبيعي للرواسب.

ز ـ قم بإعداد رسوم انحدار جزئي منفصلة مقابل $e(X_1|X_2, X_3)$، $e(X_2|X_1, X_3)$ و$e(X_3 \mid X_1, X_2)$.

ح ـ هل تقترح الرسوم في الجزئين (و)، (ز) الحاجة إلى تعديل نموذج الانحدار؟

ط ـ تقترح الحجج النظرية استخدام دالة الانحدار:

$$E\{\log_e Y\} = \beta_0 + \beta_1 \log_e X_1 + \beta_2 \log_e (140 - X_2) + \beta_3 \log_e X_3$$

هل تتسق الرسوم في الجزئين (و)، (ز) مع التوقعات النظرية؟ ناقش

(١٢-٢٠) بالإشارة إلى مسألة **وظيفة الكلية** (١٢-١٩). نريد توفيق دالة الانحدار المستندة إلى اعتبارات نظرية:

$$E\{\log_e Y\} = \beta_0 + \beta_1 \log_e X_1 + \beta_2 \log_e (140 - X_2) + \beta_3 \log_e X_3$$

وتقويمها تقويما مفصلا..

أ ـ قم بتوفيق دالة الانحدار المستندة إلى اعتبارات نظرية.

ب ـ أوجد الرواسب وارسمها بصورة منفصلة مقابل $\hat{Y}$ وكل من المتغيرات المستقلة في النموذج التوفيقي. قم أيضا بإعداد رسم احتمال طبيعي للرواسب. هل حُذفت الآن، وإلى حد كبير، الصعوبات التي لوحظت في المسألة (١٢-١٩)؟

جـ ـ أوجد عوامل تضخم التباين لنموذج الانحدار التوفيقي في الجزء (أ). هل هناك مايشير إلى وجود مشاكل جدية لخطية متعددة هنا؟ اشرح.

د ـ أوجد العناصر القطرية لمصفوفة القبعة. ومستخدما قاعدة الإبهام $2p / n$ حدد أية مشاهدات قاصية في X.

هـ ـ أوجد رواسب الحذف المعيّرة تقديرا وحدد أية مشاهدات قاصية في Y.

و ـ المشاهدتان 28 و29 قاصيتان نسبيا في قيم Y أوجد قيم *DFFITS*، *DFBETAS* ومسافة كوك لهذه المشاهدات بغية تثمين نفوذها. ماذا تستنتج؟

(١٢-٢١) بالإشارة إلى مسألتي راحة **المريض** (٧-١٧) و(١٢-٩). كان مدير المستشفى مهتما بمعرفة كيفية أداء طريقة اختيار الخطوة فخطوة إلى الأمام وبعضا من أشكالها المختلفة.

أ ـ حدد المجموعة الجزئية من المتغيرات التي تختارها، كأفضل مجموعة جزئية، بطريقة الانحدار خطوة فخطوة إلى الأمام مستخدما 3.0 و2.9 كحدين لِـ F من أجل إضافة أو حذف متغير، على الترتيب. بيّن خطواتك.

ب ـ في أي اختبار F بمفرده، ماهو، بصورة تقريبية، مستوى المعنوية المكافىء للقيمة 3.0 كحد F لإضافة متغير؟

جـ ـ حدد المجموعة الجزئية من المتغيرات التي تختارها، كأفضل مجموعة جزئية، طريقة الاختيار بالإضافة مستخدما 3.0 كحد F لإضافة متغير. بيّن خطواتك.

د ـ حدد المجموعة الجزئية من المتغيرات التي تختارها، كأفضل مجموعة جزئية، بطريقة الاختيار بالحذف، مستخدما 2.9 كحد F لحذف متغير. بيّن خطواتك.

هـ ـ قارن نتائج طرق الاختيار الثلاث. إلى أي حد تتسق هذه النتائج ؟ كيف تجد هذه النتائج بالمقارنة مع تلك الخاصة بجميع الانحدارات الممكنة في المسألة (١٢-٩)؟

(١٢-٢٢) بالعودة إلى مسألتي **كسوة السقف بالألواح** (١٢-١٠) و(١٢-١١).

أ ـ مستخدما انحدار الخطوة فخطوة إلى الأمام، أوجد المجموعة الجزئية الأفضل من المتغيرات المستقلة للتنبؤ بالمبيعات. استخدم 4.0 و 3.9 كحدي F لإضافة وحذف متغير، على الترتيب.

ب ـ كيف تجد المجموعة الجزئية الأفضل وفقا لانحدار الخطوة فخطوة إلى الأمام بالمقارنة مع أفضل مجموعة جزئية وفقا للمعيار C_p التي حصلت عليها في المسألة (١٢-١١)أ؟

جـ ـ لنفترض أننا قسرنا المتغير X_1 إلى المجموعة الجزئية الأفضل نظرا لأهميته السببية وذلك بإدخاله أولا إلى النموذج وعدم حذفه حتى لو كانت قيمة F^* من أجله متدنية جدا. ما هي المجموعة الجزئية من المتغيرات (بما في ذلك X_1) التي تختارها الآن طريقة الانحدار خطوة فخطوة إلى الأمام كأفضل مجموعة جزئية إذا اتخذنا 3.9 و 4.0 كحدي F لإضافة أو حذف متغير، على الترتيب؟ هل يؤثر الاحتواء القسري لـ X_1 في اختيار المتغيرات الأخرى للمجموعة الأفضل ؟ هل سيحدث هذا دائما ؟

(١٢-٢٣) بالعودة إلى مسألتي **البراعة في عمل** (١٢-١٣) و(١٢-١٤).

أ ـ مستخدما انحدار الخطوة فخطوة إلى الأمام، أوجد المجموعة الجزئية الأفضل من المتغيرات للتنبؤ بالبراعة في العمل، استخدم 4.0 و3.9 كحدي F لإضافة وحذف متغير، على الترتيب.

ب ـ كيف تجد المجموعة الجزئية الأفضل وفقا لانحدار خطوة فخطوة إلى الأمام بالمقارنة مع المجموعة الجزئية الأفضل وفقا لمعيار R^2 المعدل التي وجدتها في المسألة (١٢-١٤)أ؟.

(١٢-٢٤) بالعودة إلى مسألتي **كسوة السقف بالألواح** (١٢-١٠) و(١٢-١٢). لتثمين المقدرة التنبؤية لنموذج الانحدار المحدد في المسألة (١٢-١٢)، بصورة داخلية، احسب الإحصاءة *PRESS* وقارنها بـ *SSE*. ماذا تقترح هذه المقارنة حول مشروعية *SSE* كمؤشر للقدرة التنبؤية للنموذج التوفيقي؟

(١٢-٢٥) بالعودة إلى مسألتي **كسوة السقف بالألواح** (١٢-١٠) و(١٢-١٢). لتثمين صحة نموذج الانحدار المحدد في المسألة (١٢-١٢)، بصورة خارجية، عُرضت بيانات المبيعات في العام السابق لثلاث وعشرين منطقة أخرى من المناطق المشابهة. وفيما يلي هذه البيانات:

Y_i	X_{i4}	X_{i3}	X_{i2}	X_{i1}	المنطقة i
291.5	10	7	70	5.3	27
277.9	20	10	83	5.3	28
48.0	3	14	46	7.5	29
213.4	20	7	51	6.2	30
135.1	9	11	52	5.4	31
150.0	4	9	45	5.8	32
180.1	12	9	54	2.5	33
295.8	8	6	70	5.4	34
91.3	6	9	31	5.2	35
116.7	11	10	46	3.7	36
190.3	6	8	49	5.1	37
155.3	10	12	60	6.0	38
217.6	9	7	51	6.7	39
236.5	9	7	55	6.1	40
202.2	4	8	55	3.6	41
64.7	8	10	24	5.8	42
287.2	15	5	63	7.1	43
128.0	10	8	30	6.8	44
219.7	4	8	62	5.0	45
272.9	17	8	75	6.9	46
168.0	10	9	47	4.5	47
269.1	17	8	67	6.2	48
115.6	11	11	48	6.5	49

أ ـ أوجد مصفوفة الارتباط للمتغيرات X من أجل مجموعة بيانات التقويم وقارنها بتلك التي حصلت عليها في المسألة (١٢ـ١٠)جـ من أجل مجموعة بيانات بناء النموذج. ماذا تستنتج؟

ب ـ قم بتوفيق نموذج الانحدار المحدد في المسألة (١٢-١٢) لمجموعة بيانات التقويم قارن معاملات الانحدار المقدّرة وانحرافاتها المعيارية المقدرة بتلك التي حصلت عليها في المسألة (١٢-١١)أ. وقارن أيضا متوسط مربعات الخطأ ومعاملات التحديد المتعدد. هل تبدو تقديرات مجموعة بيانات التقويم مشابهة بصورة معقولة لتلك التي حصلت عليها باستخدامك لمجموعة بيانات بناء النموذج ؟

جـ ـ احسب متوسط مربعات خطأ التنبؤ في (12.23) وقارنه بـ MSE الذي حصلت عليه من مجموعة بيانات بناء النموذج. هل يبدو أن هناك مشكلة انحياز في MSE؟ هل تتسق هذه النتيجة مع ماوجدته في المسألة (١٢-٢٤)؟ علّق.

د ـ بعد ضم مجموعة بيانات بناء النموذج في المسألة (١٢ـ١٠) مع مجموعة بيانات التقويم، قم بتوفيق نموذج الانحدار المختار لمجموعة البيانات الموحدة. هل انخفضت الآن الانحرافات المعيارية المقدرة لمعاملات الانحدار المقدرة انخفاضا ملحوظا ؟

(١٢-٢٦) بالعودة إلى مسألتي **البراعة في عمل** (١٢-١٣) و(١٢-١٥). لتثمين القدرة التنبؤية لنموذج الانحدار المحدد في المسألة (١٢ـ١٥)، بصورة داخلية، احسب الإحصاءة $PRESS$ وقارنها بـ SSE. ماذا تقترح هذه المقارنة بالنسبة لمشروعية MSE كمؤشر للمقدرة التنبؤية للنموذج التوفيقي؟.

(١٢-٢٧) بالعودة إلى مسألتي **البراعة في عمل** (١٢-١٣) و(١٢-١٥). لتثمين صحة نموذج الانحدار المحدد في المسألة (١٢-١٥) و(١٢-٢٥)، بصورة خارجية، اختُبرت بصورة مماثلة لمجموعة إضافية من المتقدمين لوظائف مكتبية بسيطة

في الوكالة واستخدموا بصرف النظر عن درجاتهم. وفيما يلي البيانات الخاصة بهم.

الشخص	درجة الاختبار				درجة البراعة في العمل
i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	Y_i
26	65	109	88	84	58
27	85	90	104	98	92
28	93	73	91	82	71
29	95	57	95	85	77
30	102	139	101	92	92
31	63	101	93	84	66
32	81	129	88	76	61
33	111	102	83	72	57
34	67	98	98	84	66
35	91	111	96	84	75
36	128	99	98	89	98
37	116	103	103	103	100
38	105	102	88	83	67
39	99	132	109	105	111
40	93	95	106	98	97
41	99	113	104	95	99
42	110	114	91	78	74
43	128	134	108	98	117
44	99	110	96	97	92
45	111	113	101	91	95
46	109	120	104	106	104
47	78	125	115	102	100
48	115	119	102	94	95
49	129	70	94	95	81
50	136	104	106	104	109

أ - أوجد مصفوفة ارتباط المتغيرات X لمجموعة بيانات التقويم وقارنها بتلك التي حصلت عليها في المسألة (١٢-١٣)جـ لمجموعة بيانات بناء النموذج. هل مصفوفتا الارتباط متماثلتان بصورة معقولة ؟

ب - قم بتوفيق نموذج الانحدار المحدد في المسألة (١٢-١٥) لمجموعة بيانات التقويم، قارن معاملات الانحدار المقدرة وانحرافاتها المعيارية المقدَّرة بتلك التي حصلت عليها في المسألة (١٢-١٤)أ. قارن أيضا متوسط

مربعات الخطأ ومعاملات التحديد المتعدد. هل تبدو تقديرات بيانات التقويم مماثلة بصورة معقولة لتقديرات بيانات بناء النموذج ؟

جـ - احسب متوسط مربعات خطأ التنبؤ (12.23) وقارنها بـ MSE الناتجة عن بيانات بناء النموذج. هل هناك دلالة على وجود مشكلة انحياز شديد في MSE هنا ؟ لعل هذه النتيجة متسقة مع ماوجدته في المسألة (١٢-٢٦)؟ ناقش.

د - بعد ضم مجموعة بيانات بناء النموذج في المسألة (١٢-١٣) مع مجموعة بيانات التقويم في مجموعة واحدة، قم بتوفيق نموذج الانحدار المختار إلى المجموعة الموحدة. هل انخفضت الآن الانحرافات المعيارية المقدرة لمعاملات الانحدار المقدّرة انخفاضا ملحوظا عما كانت عليه عند استخدام مجموعة بيانات بناء النموذج؟.

(١٢-٢٨) بالعودة إلى مسألتي **ضغط الرئة** (١٢-١٦) و(١٢-١٨). نريد تثمين صحة نموذج الانحدار المحدد في المسألة (١٢-١٨)، بصورة داخلية.

أ - احسب الإحصاءة $PRESS$ وقارنها بـ SSE. ماذا تقترح هذه المقارنة بالنسبة لمشروعية MSE كمؤشر للقدرة التنبؤية للنموذج التوفيقي؟

ب - تتسبب المشاهدة 8 بمفردها بنصف القيمة الإجمالية لإحصاءة $PRESS$ تقريبا. هل توصي بتعديل النموذج بسبب التأثير القوي لهذه المشاهدة؟ ماهي اختيارات العمل التصحيحي التي يمكن أن تخفف من تأثير المشاهدة 8 ؟ ناقش.

تمارين

(١٢-٢٩) دالة الانحدار التربيعية الصحيحة هي $E\{Y\} = 15+20X+3X^2$ ودالة الانحدار التوفيقية الخطية هي $\hat{Y}=13+40X$ ، ومن أجل هذه الدالة $E\{b_0\} = 10$ و$E\{b_1\} = 45$، ما هما مركبتا الانحياز وخطأ المعاينة لمتوسط مربعات الخطأ لـ $X_i = 10$ ولـ $X_i = 20$؟

(١٢-٣٠) برهن (12.12) إرشاد : استخدم التمرين ٦-٣١ و (11.14)

(١٢-٣١) بالإشارة إلى (12.18) بيّن أن المتغير X_l الذي يجعـل إحصـاءة الاختبـار F_k^* أعظم ما يمكن هو نفسه يجعل معامل التحديد الجزئي $r_{Yk.?}^2$ أعظـم مـايمكن أيضا.

مشاريع

(١٢-٣٢) بالعودة إلى مجموعة بيانات بناء النموذج في مثال وحدة الجراحة في الجـدول (١٢-١) ومجموعـة بيانـات التقويـم في الجـدول (١٢-٦). ضـم مجموعـتي البيانات في مجموعة واحدة وقم بتوفيق نموذج الانحدار (12.21) للمجموعة الموحدة من البيانات. قارن معاملات الانحدار المقـدرة وانحرافاتهـا المعياريـة المقدرة بتلك التي حصلت عليها في الجدول (١٢-٥) مـن مجموعـة بيانـات بنـاء النمـوذج. هـل هنـاك فـروق كبـيرة في الانحرافـات المعياريـة المقـدّرة لمعاملات الانحدار المقدّرة ؟ علّق.

(١٢-٣٣) بالعودة إلى مجموعة البيانات **SENIC** نريد التنبؤ بطول الإقامة Y، وتتضمـن جملة المتغيرات المستقلة المرشحة جميع المتغيرات الأخرى في مجموعة البيانات باستثناء الانتمـاء إلى مدرسـة طـب والمنطقـة. ويُعتقـد أن نموذجـا يتضمـن $log_{10}Y$ كمتغير تابع ومتغـيرات مسـتقلة بحـدود مـن المرتبـة الأولى وبـدون حدود تفـاعل سـيكون نموذجـا مناسـبا. لنـأخذ الحـالات مـن 57 إلى 113 لتشكل مجموعة بيانات بناء النموذج ولاستخدامها في التحليلات التالية:

أ - قم بإعداد رسوم نقطية منفصلة لكل من المتغيرات المستقلة. هل هناك أية نواح تستحق الملاحظة في هذه الرسوم؟ علّق.

ب - أوجد مصفوفة الارتباط للمتغيرات X. هل هنـاك دليـل علـى وجـود صلات خطية بين أزواج المتغيرات؟

جـ - أوجد المجموعات الجزئية الثلاث الأفضل وفقا للمعيار C_p أيّ النماذج الجزئية هذه يبدو أقل انحيازا؟

(١٢-٣٤) بالعودة إلى مجموعة بيانات *SENIC* والمسألة (١٢-٣٣). نريد تقويما تفصيليا لنموذج الانحدار المتضمن للعمر ونسبة التصوير الشعاعي الروتيني للصدر، ومتوسط التعداد اليومي. في حدود من المرتبة الأولى، وذلك بالاستناد إلى مجموعة بيانات بناء النموذج.

أ - أوجد الرواسب وارسمها بصورة منفصلة مقابل $\hat{Y}$، ومقابل كل من المتغيرات المستقلة في النموذج، وكل من الحدود الجدائية ذات الصلة على أساس من هذه الرسوم هل ينبغي القيام بأية تعديلات في النموذج؟

ب - قم بإعداد رسم احتمال طبيعي للرواسب. وأوجد أيضا معامل الارتباط بين الرواسب المرتبة وقيمها المتوقعة تحت الطبيعية. اختبر معقولية افتراض الطبيعية مستخدما الجدول (٤-٣) و$\alpha = 0.05$ ماذا تستنتج؟

جـ - أوجد عوامل تضخم التباين. هل هناك مؤشرات لوجود مشاكل جدية لخطية متعددة؟ اشرح.

د - أوجد العناصر القطرية لمصفوفة القبعة. ومستخدما قاعدة الباهم $2p/n$ حدد أية مشاهدات قاصية في X.

هـ - أوجد رواسب الحذف المعيرة تقديرا وحدد المشاهدات القاصية في Y.

و - المشاهدات 62، 75، 106 و112 قاصية بصورة معتدلة بالنسبة لقيم X، والمشاهدة 87 قاصية إلى حد ما بالنسبة لقيمة Y أوجد قيم *DFBETAS*، *DFFITS* ومسافة كوك لهذه المشاهدات لتثمين نفوذها. ماذا تستنتج؟

ز - احسب الإحصاءة *PRESS* وقارنها بـ *SSE* ماذا تقترح هذه المقارنة بالنسبة لصلاحية *MSE* كمؤشر للقدرة التنبؤية للنموذج التوفيقي؟

(١٢-٣٥) بالعودة إلى مجموعة بيانات **SENIC** والمسألتين (١٢-٣٣) و(١٢-٣٤). نريد التحقق من صحة نموذج الانحدار المحدد في المسألة (١٢-٣٤) بواسطة مجموعة بيانات التقويم المؤلفة من المشاهدات 1 إلى 56.

أ ـ قم بتوفيق نموذج الانحدار المحدد في المسألة (١٢-٣٤) لمجموعة بيانات التقويم وقارن معاملات الانحدار المقدرة وانحرافاتها المعيارية المقدرة بتلك التي حصلت عليها في المسألة (١٢-٣٣)جـ. قارن أيضا متوسطي مربعات الخطأ ومعاملي التحديد المتعدد. هل ينتج النموذج التوفيقي لمجموعة بيانات التقويم تقديرات مشابهة لتقديرات النموذج التوفيقي لمجموعة بيانات بناء النموذج ؟

ب ـ احسب متوسط مربعات خطأ التنبؤ في (12.23) وقارنه بـ MSE الذي حصلت عليه من مجموعة بيانات بناء النموذج. هل هناك دليل على وجود انحياز شديد في MSE هنا؟ هل تتسق هذه النتيجة مع ما وجدته في المسألة (١٢-٣٤)؟.

جـ ـ ضم مجموعتي بيانات التقويم وبناء النموذج في مجموعة واحدة ثم قم بتوفيق نموذج الانحدار المختار لمجموعة البيانات الموحدة. هل تختلف معاملات الانحدار المقدّرة وانحرافاتها المعيارية المقدرة اختلافا كبيرا عن تلك التي وجدتها من مجموعة بيانات بناء النموذج؟ هل ينبغي توقع أية فروق ؟ اشرح

(١٢-٣٦) بالاشارة إلى مجموعة بيانات **SMSA** يرغب مسؤول سلامة عامة بالتنبؤ بمعدل الجرائم في **SMSA** (Y، العدد الكلي للجرائم الخطرة لكل مائة ألف من السكان). وتتضمن جملة المتغيرات المستقلة المرشحة جميع المتغيرات الأخرى في مجموعة البيانات باستثناء العدد الإجمالي للسكان والمنطقة. ويُعتقد أن النموذج المتضمن لمتغيرات مستقلة في حدود من المرتبة الأولى بدون حدود تفاعل سيكون نموذجا مناسبا. خذ المشاهدات ذات الأرقام المتسلسلة الزوجية لتشكل مجموعة بيانات بناء النموذج ولاستخدامها في التحليلات التالية.

أ ـ قم باعداد رسوم جذع وورقة لكل من المتغيرات المستقلة. هل هناك أية نواحٍ تستحق الملاحظة في هذه الرسوم؟ علّق.

ب ـ أوجد مصفوفة ارتباط المتغيرات X. هل هناك دليل على وجود صلات خطية قوية بين أزواج من المتغيرات هنا؟

جـ ـ مستخدما المعيار C_p، أوجد أفضل ثلاث مجموعات جزئية. أي هذه النماذج يبدو أنه النموذج ذو الانحياز الأقل؟

(١٢-٣٧) بالعودة إلى مجموعة بيانات ***SMSA*** والمسألة (١٢-٣٦). نريد تقويما تفصيليا لنموذج الانحدار المتضمن لمساحة المنطقة والنسبة المئوية للحاصلين على الثانوية العامة في حدود من المرتبة الأولى، وذلك بالاستناد إلى مجموعة بيانات بناء النموذج.

أ ـ أوجد الرواسب وارسمها بصورة منفصلة مقابل $\hat{Y}$، ومقابل كل من المتغيرات المستقلة في النموذج، والحد الجدائي المتعلق بها. وعلى أساس من هذه الرسوم هل ينبغي القيام بأية تعديلات في النموذج؟

ب ـ قم بإعداد رسم احتمال طبيعي للرواسب. أوجد أيضا معامل الارتباط بين الرواسب المرتبة وبين قيمها المتوقعة تحت الطبيعية. اختبر معقولية افتراض الطبيعية مستخدما الجدول $\alpha = 0.01$. ماذا تستنتج؟

جـ ـ أوجد عوامل تضخم التباين. هل هناك أية مؤشرات على وجود مشاكل حدية لخطية متعددة ؟ اشرح.

د ـ أوجد العناصر القطرية لمصفوفة القبعة. ومستخدما قاعدة الإبهام $2\,p\,/\,n$ حدد المشاهدات القاصية في X.

هـ ـ أوجد رواسب الحذف المعيرة تقديرا وحدد المشاهدات القاصية في Y.

و ـ المشاهدات 42، 74، 92، 94، 124 و138 قاصية بصورة معتدلة بالنسبة لقيم X، والمشاهدتان 40 و54 قاصيتان إلى حد ما بالنسبة

لقيم Y. احسب قيم *DFFITS*، *DFBETAS* ومسافة كوك لهذه المشاهدات لتثمين نفوذها. ماذا تستنتج؟

ز ـ احسب الإحصاءة *PRESS* وقارنها بـ *SSE*. ماذا تقترح هذه المقارنة بشأن صلاحية *MSE* كمؤشر للقدرة التنبؤية للنموذج التوفيقي؟

(١٢ـ٣٨) بالعودة إلى مجموعة بيانات ***SMSA*** والمسألتين (١٢ـ٣٦) و(١٢ـ٣٧) نريد التحقق من صحة نموذج الانحدار المحدد في المسألة (١٢ـ٣٧) مستخدمين مجموعة بيانات التقويم المؤلفة من المشاهدات ذات الأرقام المتسلسلة الفردية في مجموعة البيانات.

أ ـ قم بتوفيق نموذج الانحدار المحدد في المسألة (١٢-٣٧) إلى مجموعة بيانات التقويم. قارن معاملات الانحدار المقدرة وانحرافاتها المعيارية المقدرة بتلك التي حصلت عليها في المسألة (١٢ـ٣٦)جـ. قارن أيضا متوسطي مربعات الخطأ ومعاملي التحديد المتعدد. هل ينتج النموذج التوفيقي لمجموعة بيانات التقويم تقديرات مشابهة لتلك الخاصة بالنموذج التوفيقي لمجموعة بيانات بناء النموذج؟

ب ـ احسب متوسط مربعات خطأ التنبؤ في (12.23) وقارنه بـ *MSE* الذي حصلت عليه من مجموعة بيانات بناء النموذج. هل هناك دليل على وجود انحياز شديد في *MSE* هنا؟ هل تتسق هذه النتيجة مع ماوجدته في المسألة (١٢ـ٣٧)و.

جـ ـ قم بتوفيق نموذج الانحدار المحدد في المسألة (١٢ـ٣٧) لمجموعة البيانات الناتجة عن ضم مجموعتي بيانات التقويم وبيانات بناء النموذج. هل تختلف معاملات الانحدار المقدرة وانحرافاتها المعيارية المقدرة اختلافا واضحا عن تلك الخاصة بالنموذج التوفيقي لمجموعة بيانات بناء النموذج؟ هل ينبغي لك توقع أية فروق في التقديرات؟ اشرح.

الارتباط الذاتي في بيانات السلاسل الزمنية

افترضت نماذج الانحدار الأساسية التي درسناها حتى الآن أن حدود الخطأ العشوائي ε_t إما أن تكون متغيرات عشوائية غير مرتبطة أو متغيرات عشوائية مستقلة تتبع التوزيع الطبيعي. وفي الدراسات الاقتصادية والتجارية هناك الكثير من تطبيقات الانحدار التي تتضمن بيانات سلاسل زمنية.

وافتراض عدم ارتباط أو استقلال حدود الخطأ في مثل هذه البيانات لايكون في الغالب مناسبا، بل كثيرا ماتكون حدود الخطأ مرتبطة ارتباطا موجبا فوق الزمن. وحذف واحد أو أكثر من المتغيرات الأساسية من نموذج الانحدار هو سبب رئيس لنشوء ارتباط ذاتي موجب بين حدود الخطأ في تطبيقات نماذج الانحدار في الدراسات الاقتصادية والتجارية التي تتضمن بيانات سلاسل زمنية. وعندما تكون التأثيرات المرتبة زمنيا للمتغيرات المحذوفة مرتبطة ارتباطا موجبا تنحو حدود الخطأ في نموذج الانحدار إلى أن يكون ارتباطها الذاتي موجبا أيضا وذلك لأن حدود الخطأ تتضمن تأثيرات المتغيرات المحذوفة، افترض مثلا أننا حدرنا المبيعات السنوية لمنتج ما على متوسط السعر السنوي لمدة 30 سنة. فإذا كان لحجم المجتمع أثر مهم على المبيعات فقد يؤدي حذفه من النموذج إلى أن تكون حدود الخطأ مرتبطة ارتباطا ذاتيا موجبا، إذ يتوقع أن تكون تأثيرات حجم المجتمع على المبيعات مرتبطة ارتباطا موجبا فوق الزمن.

والسبب الشائع الآخر لظهور ارتباط ذاتي موجب في حدود الخطأ في البيانات الإقتصادية هو أخطاء تغطية نمطية في السلاسل الزمنية للمتغير التابع، فغالبا ما تكون مثل هذه الأخطاء مرتبطة ارتباطا موجبا فوق الزمن.

(١٣-١) مشاكل الارتباط الذاتي

إذا كانت حدود الخطأ في نموذج الانحدار مرتبطة ارتباطا ذاتيا موجبا، فإن استخدام طرق المربعات الدنيا العادية يترتب عليه عدد من العواقب المهمة. وسنلخص هذه العواقب أولا ومن ثم نناقشها بتفصيل أكثر:

١- لاتزال معاملات الانحدار المقدَّرة غير منحازة إلا أنها لاتملك الآن خاصية التباين الأصغر ويمكن أن تكون غير فعّالة.

٢- متوسط مربعات الخطأ يمكن أن يشكل تقديرا بالنقصان لتباين حدود الخطأ.

٣- قد تُعطي $s\{b_k\}$ محسوبة وفقا لطرق المربعات الدنيا العادية، تقديرا بالنقصان للانحراف المعياري الحقيقي لمعامل الانحدار المقدَّر.

٤- لم تعد فترات الثقة والاختبارات التي تستخدم توزيعات t وF. والتي سبق مناقشتها قابلة للتطبيق.

ولتوضيح هذه المشاكل اعتمادا على البداهة، سوف ندرس نموذج الانحدار الخطّي البسيط مع بيانات سلاسل زمنية.

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

هنا Y_t وX_t هي مشاهدات في الفترة t. لنفترض أن حدود الخطأ مرتبطة ذاتيا بصورة موجبة كما يلي:

$$\varepsilon_t = \varepsilon_{t-1} + u_t$$

وتُسمى المتغيرات العشوائية u_t، "الاضطرابات"، وهي تغيرات عشوائية طبيعية مستقلة. وهكذا فإن أي حد خطأ ε_t هو مجموع حد الخطأ السابق ε_{t-1} وحد اضطراب جديد u_t. وسنفترض هنا أن u_t له متوسط 0 وتباين 1 .

وفي العمود الأول من الجدول (١٣-١) نعطي 10 مشاهدات عشوائية للمتغير الطبيعي u_t بمتوسط 0 وتباين 1، تم الحصول عليها من مولد لأعداد عشوائية طبيعية معيارية. لنفرض الآن أن $\varepsilon_0 = 3.0$ فعندئذ نجد:

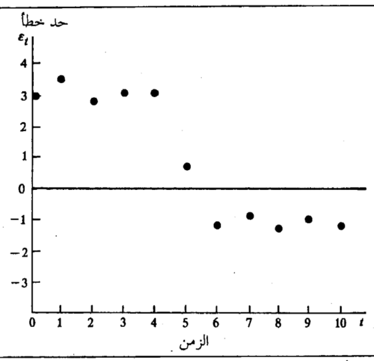
$$\varepsilon_1 = \varepsilon_0 + u_1 = 3.0 + 0.5 = 3.5$$
$$\varepsilon_2 = \varepsilon_1 + u_2 = 3.5 - 0.7 = 2.8$$
$$etc.$$

وحدود الخطأ مبينة في العمود الثاني من الجدول (١٣-١)، وقد تمَّ رسمها في الشكل (١٣-١). لاحظ النمط المنهجي في حدود الخطأ هذه. وتتضح علاقتها الموجبة فوق الزمن من حقيقة أن حدود الخطأ المتجاورة تتجه إلى أن تكون من الحجم نفسه.

جدول (١٣-١) مثال على حدود خطأ ترتبط ارتباطا ذاتيا موجبا.

(٢) $\varepsilon_{t-1} + u_t = \varepsilon_t$	(١) u_t	t
3.0	-	0
3.0 + .5 = 2.8	+.5	1
3.5 − .7 = 2.8	−.7	2
2.8 + .3 = 3.1	+.3	3
3.1 + 0 = 3.1	0	4
3.1 −2.3 = .8	−2.3	5
.8 − 1.9 = -1.1	−1.9	6
-1.1 + .2 = -.9	+.2	7
− .9 − .3 =−1.2	−.3	8
−1.2 + .2 = −1.0	+.2	9
−1.0 − .1 = −1.1	−.1	10

شكل (١٣-١) مثال على حدود خطأ ترتبط ارتباطا ذاتيا موجبا.

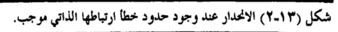


افترض أن X_t في نموذج الانحدار يمثل الزمن بحيث يكون $X_1 = 1$ و$X_2 = 2$ الخ. وبالإضافة إلى ذلك افترض أننا نعلم أن $\beta_0 = 2$ و$\beta_1 = 0.5$ بحيث تكون دالة الانحدار الحقيقية $E\{Y\} = 2 + 0.5X$. فيوضح الشكل (١٣-٢)أ. خط الانحدار الحقيقي وقيم Y الملحوظة والمبنية على حدود الخطأ في الجدول (١٣-١). وعلى سبيل المثال:

$Y_0 = 2 + 0.5\ (0) + 3.0 = 5.0$ و $Y_1 = 2 + 0.5(1) + 3.5 = 6.0$، ويوضح الشكل (١٣-٢)ب. خط الانحدار المقدَّر والذي تم توفيقه بطرق المربعات الدنيا العادية مع إعادة رسم قيم Y الملحوظة. لاحظ أن خط الانحدار التوفيقي يختلف بشكل حاد عن خط الانحدار الحقيقي، ذلك لأن القيمة الابتدائية ε_0 كانت كبيرة، وأن القيم اللاحقة لها مرتبطة ارتباطا ذاتيا موجبا، مما تسبب في أن تبقى قيمها كبيرة لبعض الوقت. ويقود هذا النمط المتواصل لحدود الخطأ ذات الارتباط الذاتي الموجب إلى خط انحدار توفيقي بعيد كل البعد عن الخط الحقيقي. ولو كانت القيمة الابتدائية ε_0 صغيرة ولتكن $\varepsilon_0 = -0.2$ والاضطرابات مختلفة، فربما حصلنا على خط انحدار توفيقي مختلف تماما وذلك بسبب نمط التواصل كما هو موضح في الشكل (١٣-٢)جـ. وهذا التغير من عينة إلى عينة في خطوط الانحدار التوفيقية والذي يعود إلى حدود الخطأ ذات الارتباط الذاتي الموجب، يمكن أن يكون كبيرا إلى الحد الذي يقود إلى تباينات كبيرة لمعاملات الانحدار المقدّرة وذلك عند استخدام طرق المربعات الدنيا العادية.

واستخدام طرق المربعات الدنيا العادية عندما تكون حدود الخطأ ذات ارتباط ذاتي موجب، يقود إلى مشكلة أساسية أخرى هي، كما ذكرنا سابقا، أن MSE، متوسط مربعات الخطأ، يمكن أن يعطي تقديرا بالنقصان لتباين الحدود. ويوضح الشكل (١٣-٢) هذه النقطة. لاحظ أن تشتيت قيم Y حول خط الانحدار التوفيقي في الشكل (١٣-٢)ب أصغر بكثير من تشتت هذه القيم حول خط الانحدار الحقيقي في الشكل (١٣-٢)أ؛ وهذا هو أحد العوامل التي تشير إلى إحكام أكبر في معاملات الانحدار أكبر مما هو عليه الحال، في الحقيقة، عند استخدام فرق المربعات الدنيا العادية، مع وجود أخطاء ارتباطها الذاتي موجب.

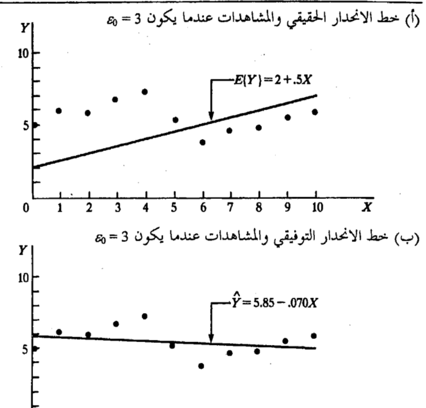
وفي ضوء خطورة المشاكل الناشئة عن وجود أخطاء ذاتية الارتباط فإن التحقق من وجودها هو من الأهمية بمكان. ورسم الرواسب في مقابل الزمن هو، بالرغم من أنه لايتصف بالموضوعية، وسيلة فعالة للكشف عن وجود أخطاء ذاتية الارتباط. وقد تم أيضا تطوير اختبارات إحصائية لهذا، ويستند أحد هذه الاختبارات، وهو يُستخدم على نطاق واسع، إلى نموذج خطأ ذاتي الانحدار من المرتبة الأولى، مما نعرض له في الفقرة التالية. وهذا النموذج بسيط، علاوة على أن الخبرة تقترح قابليته المتواترة للتطبيق في الاقتصاد والتجارة، وذلك عندما تكون حدود الخطأ مرتبطة ارتباطا تسلسليا.



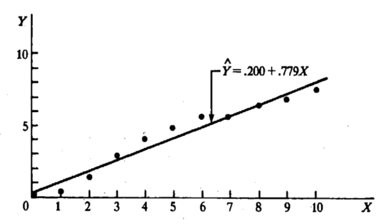
شكل (٢-١٣) الانحدار عند وجود حدود خطأ ارتباطها الذاتي موجب.

(أ) خط الانحدار الحقيقي والمشاهدات عندما يكون $\varepsilon_0 = 3$

(ب) خط الانحدار التوفيقي والمشاهدات عندما يكون $\varepsilon_0 = 3$



(جـ) خط الانحدار التوفيقي والمشاهدات عندما يكون $\varepsilon_0 = -0.2$ والاضطرابات مختلفة.



(١٣-٢) نموذج خطأ ذاتي الانحدار من المرتبة الأولى

انحدار خطي بسيط

نموذج الانحدار الخطي البسيط بمتغير مستقل واحد وحدود خطأ عشوائية تتبع نمطا ذاتي الانحدار من المرتبة الأولى هو:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \varepsilon_t \\ \varepsilon_t &= \rho\varepsilon_{t-1} + \mu_t \end{aligned} \tag{13.1}$$

حيث: ρ هي معلمة بحيث $|\rho| < 1$

u_t مستقلة و$N(0, \sigma^2)$

لاحظ أن (13.1) متطابقة مع نموذج الانحدار الخطي البسيط (3.1) باستثناء مايتعلق بتركيب حدود الخطأ. ويتألف كل حد خطأ في (13.1) من كسر من حد الخطأ السابق (عندما $\rho > 0$) بالإضافة إلى حد اضطراب جديد u_t. وتسمى المعلمة ρ معلمة الارتباط الذاتي.

انحدار متعدد

نموذج الانحدار المتعدد بحدود خطأ عشوائية تتبع نمطا ذاتي الانحدار من المرتبة الأولى هو:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_{p-1} X_{t,p-1} + \varepsilon_t \\ \varepsilon_t &= \rho\varepsilon_{t-1} + \mu_t \end{aligned} \tag{13.2}$$

حيث:

$|\rho| < 1$

u_t مستقلة و$N(0, \sigma^2)$

وهكذا نرى أن نموذج الانحدار المتعدد (13.2) متطابق مع نموذج الانحدار المذكور سابقا (7.7) باستثناء مايتعلق بتركيب حدود الخطأ.

خواص حدود الخطأ

ومن المفيد تعميم تعريف حد الخطأ ذاتي الانحدار من المرتبة الأولى، ε_t كما يلي:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

بحيث يبقى صحيحا لكل قيم t وبالتالي يكون $\varepsilon_{t-1} = \rho\varepsilon_{t-2} + u_{t-1}$ وعند تعويض هذه العبارة أعلاه، نحصل على:

$$\varepsilon_t = \rho(\rho\varepsilon_{t-2} + u_{t-1}) + u_t = \rho^2\varepsilon_{t-2} + \rho u_{t-1} + u_t$$

والآن بوضع $\rho\varepsilon_{t-3} + u_{t-2}$ مكان ε_{t-2} نحصل على:

$$\varepsilon_t = \rho^3\varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t$$

وبالاستمرار بهذه الطريقة نجد:

$$\varepsilon_t = \sum_{s=0}^{\infty} \rho^s u_{t-s} \tag{13.3}$$

وهكذا يكون حد الخطأ ε_t في الفترة t تركيبا خطيا في حد الاضطراب الراهن والحدود السابقة له. وعندما يكون $0 < \rho < 1$، فإن (13.3) تشير إلى أنه كلما بعدت الفترة في الماضي كلما كان لحد الاضطراب وزن أقل في تحديد قيمة ε_t.

ويمكن تبيان أن المتوسط والتباين لـε_t في نماذج خطأ الانحدار الذاتي من المرتبة الأولى(13.1) و (13.2) هي كالتالي:

$$E\{\varepsilon_t\} = 0 \tag{13.4}$$

$$\sigma^2\{\varepsilon_t\} = \frac{\sigma^2}{1-\rho^2} \tag{13.5}$$

وهكذا نرى أن لحدود الخطأ ε_t متوسطا يساوي الصفر وتباينا ثابتا تماما كما في حالة نماذج الانحدار بحدود أخطاء غير مرتبطة.

وخلافا، لنماذج الانحدار السابقة، على أي حال، فإن حدود الخطأ مرتبطة في نماذج خطأ الانحدار الذاتي من المرتبة الأولى(13.1) و(13.2). ويمكن إثبات أن التغاير بين حدود الخطأ المتجاورة ε_{t-1} و ε_t هو:

$$\sigma\{\varepsilon_t, \varepsilon_{t-1}\} = \rho\left(\frac{\sigma^2}{1-\rho^2}\right) \tag{13.6}$$

ويُعرَّف معامل الارتباط بين ε_{t-1} و ε_t ويرمز له بـ $\rho\{\varepsilon_t, \varepsilon_{t-1}\}$ كما يلي:

$$\rho\{\varepsilon_t, \varepsilon_{t-1}\} = \frac{\sigma\{\varepsilon_t, \varepsilon_{t-1}\}}{\sigma\{\varepsilon_t\}\sigma\{\varepsilon_{t-1}\}} \tag{13.7}$$

وبما أن التباين لكل حد من حدود الخطأ وفقا لِ (13.5) هو $\sigma^2/(1-\rho^2)$ فإن معامل الارتباط هو:

$$\rho\{\varepsilon_t, \varepsilon_{t-1}\} = \frac{\rho\left(\frac{\sigma^2}{1-\rho^2}\right)}{\sqrt{\frac{\sigma^2}{1-\rho^2}}\sqrt{\frac{\sigma^2}{1-\rho^2}}} = \rho \qquad (13.7a)$$

وهكذا فإن معلمة الارتباط الذاتي ρ هي نفسها معامل الارتباط بين حدود الأخطاء المتجاورة.

ويمكن تبيان أن التغاير بين حدي الخطأ اللذين يفصل بينهما s من الفترات هو:

$$\sigma\{\varepsilon_t, \varepsilon_{t-s}\} = \rho^s\left(\frac{\sigma^2}{1-\rho^2}\right) \qquad s \neq 0 \qquad (13.8)$$

ولذلك فإن معامل الارتباط بين ε_t وε_{t-s} هو:

$$\rho\{\varepsilon_t, \varepsilon_{t-s}\} = \rho^s \qquad s \neq 0 \qquad (13.9)$$

ومن ذلك يتضح لنا أنه إذا كانت ρ موجبة فإن كل حدود الخطأ تكون مرتبطة ولكن كلما تباعدت الحدود كلما قل ارتباطها. والحالة الوحيدة التي تكون فيها حدود الخطأ في نماذج الخطأ ذاتي الانحدار (13.1) و(13.2) غير مرتبطة هي عندما يكون $\rho = 0$.

تعليقات

١- استنباط (13.4) ، أي أن لحدود الخطأ توقعا يساوي الصفر، يتبع مباشرة من أخذ توقع ε_t في (13.3) واستخدام $E\{u_t\} = 0$ لكل t وفقا للنماذج (13.1) و(13.2).

٢- لاستنباط تباين حدود الخطأ، نستفيد من افتراض أن u_t مستقلة ولها تباين σ^2 في النماذج (13.1) و(13.2). وعندئذ نجد من (13.3):

$$\sigma^2\{\varepsilon_t\} = \sum_{s=0}^{\infty}\rho^{2s}\sigma^2(u_{t-s}) = \sigma^2\sum_{s=0}^{\infty}\rho^{2s}$$

ومن المعروف الآن أنه لكل $|\rho| < 1$ لدينا:

$$\sum_{s=0}^{\infty}\rho^{2s} = \frac{1}{1-\rho^2}$$

وبالتالي يكون:

$$\sigma^2\{\varepsilon_t\} = \frac{\sigma^2}{1-\rho^2}$$

٣- ولاستنباط التغاير بين ε_t وε_{t-1} تنبغي معرفة أن:

$$\sigma^2\{\varepsilon_t\} = E\{\varepsilon_t^2\}$$
$$\sigma\{\varepsilon_t, \varepsilon_{t-1}\} = E\{\varepsilon_t\varepsilon_{t-1}\}$$

وهاتان النتيجتان تتبعان من النظريتين $(1.15a)$ و$(1.21a)$ على الترتيب باعتبار أن $E\{\varepsilon_t\} = 0$ لكل قيم t وذلك من (13.4).

ومن (13.3) نجد:

$$E\{\varepsilon_t\varepsilon_{t-1}\} = E\{(u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \dots)(u_{t-1} + \rho u_{t-2} + \rho^2 u_{t-3} + \dots)\}$$

ويمكن إعادة كتابة هذه على الشكل:

$$E\{\varepsilon_t\varepsilon_{t-1}\} = E\left\{\left[u_t + \rho(u_{t-1} + \rho u_{t-2} + \dots)\right]\left[u_{t-1} + \rho u_{t-2} + \rho^2 u_{t-3} + \dots\right]\right\}$$
$$= E\left\{u_t\,(u_{t-1} + \rho u_{t-2} + \rho^2 u_{t-3} + \dots)\right\}$$
$$+ E\left\{\rho(u_{t-1} + \rho u_{t-2} + \rho^2 u_{t-3} + \dots)^2\right\}$$

وبما أن $E\{u_t\ u_{t-s}\} = 0$ لكل قيم $s \neq 0$، كنتيجة لفرضية استقلال u_t، ولكن $E\{u_t\}=0$ مهما يكن t، فيتلاشى الحد الأول ونحصل على

$$E\{\varepsilon_t\varepsilon_{t-1}\} = \rho E\{\varepsilon^2_{t-1}\} = \rho\sigma^2\{\varepsilon_{t-1}\}$$

وبالتالي نجد من (13.5) التي تصح أيا كانت قيمة t:

$$\sigma\{\varepsilon_t, \varepsilon_{t-1}\} = \rho\left(\frac{\sigma^2}{1-\rho^2}\right)$$

٤ـ تُعتبر عملية خطأ الانحدار الذاتي من المرتبة الأولى في النماذج (13.1) و(13.2) من أبسط الأنواع، وتتخذ العملية من المرتبة الثانية الشكل التالي:

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + u_t \qquad (13.10)$$

ويبقى من الممكن افتراض عمليات من مراتب أعلى. لقد تم تطوير طرق خاصة لعمليات الخطأ ذاتي الانحدار المعقدة. وقد نوقشت هذه الطرق في معالجات لعمليات السلاسل الزمنية والتنبؤ كما في المرجع [13.1] مثلا.

(١٣-٣) اختبار دربن ـ واتسون للارتباط الذاتي

يفترض اختبار دربن ـ واتسون للارتباط الذاتي نماذج الخطأ ذاتي الارتباط (13.1) أو (13.2) مع تثبيت قيم المتغير أو المتغيرات المستقلة. ويتضمن الاختبار تحديد ما إذا كانت معلمة الارتباط ρ في (13.1) أو(13.2) صفرا أم لا. لاحظ أنه إذا كان $\rho = 0$ فإن $\varepsilon_t = u_t$. وعندئذ تكون حدود الخطأ ε_t مستقلة باعتبار أن u_t مستقلة.

وبما أن حدود الخطأ في تطبيقات الاقتصاد والأعمال تنحو إلى إظهار ارتباط تسلسلي موجب فإن بديلي الاختبار المعتاد هما:

$$H_0: \rho = 0 \qquad (13.11)$$
$$H_a: \rho > 0$$

ويمكن الحصول على إحصاءة الاختبار D باستخدام طريقة المربعات الدنيا العادية لتوفيق دالة الانحدار، فنحسب أولا الرواسب العادية:

$$e_t = Y_t - \hat{Y}_t \qquad (13.12)$$

ثم نحسب الإحصاءة:

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \qquad (13.13)$$

حيث n عدد المشاهدات.

لاتوجد طريقة اختبار دقيقة ولكن دربن وواتسون حصلا على حد أدنى d_L وحد أعلى d_U بحيث إن أي قيمة لِـ D خارج نطاق هذين الحدين تقود إلى قرار حاسم. وقاعدة القرار للاختيار بين البديلين في (13.11) هي:

إذا كان : $D > d_U$ نستنتج H_0

إذا كان : $D < d_L$ نستنتج H_a (13.14)

إذا كان : $d_L \le D \le d_U$ فلا يقدم الاختبار نتائج محددة.

وتقود قيم D الصغيرة إلى النتيجة $\rho > 0$، وذلك لأن حدود الخطأ المتجاورة ε_t و ε_{t-1} تتجه عندما يكون ارتباطها الذاتي موجبا، إلى أن يكون لها الحجم نفسه. ولذلك فإن الفروقات في الرواسب $e_t - e_{t-1}$ تتجه لأن تكون قيمها صغيرة عندما تكون $\rho > 0$، وهذا يؤدي إلى أن يكون البسط في D صغيرا وبالتالي يؤدي إلى قيمة صغيرة لإحصاءة الاختبار D.

ويتضمن الجدول (أ - ٦) الحدين d_L و d_U لحجوم عينات مختلفة n وذلك من أجل مستويي معنوية 0.01 و 0.05، ولأعداد مختلفة $p - 1$ من المتغيرات المستقلة X في نموذج الانحدار.

مثال

ترغب شركة بليسدل (*Blaisdell*) التنبؤ بحجم مبيعاتها وذلك باستخدام مبيعات الصناعة كمتغير تنبؤ. (يمكن الحصول على تنبؤات دقيقة لمبيعات الصناعة من الرابطة التجارية للصناعة المعنية) وتحوي الأعمدة (١) و(٢) في الجدول (١٣-٢) البيانات ربع السنوية المعدلة فصليا لمبيعات الشركة ومبيعات الصناعة، على التوالي، وذلك للفترة ١٩٨٣-١٩٨٧م. وقد اقترح رسم انتشار (غير موضح هنا) أنه من المناسب استخدام نموذج انحدار خطي. ولكن محلل أبحاث السوق كان مهتما، على أي حال، فيما إذا كان الارتباط الذاتي لحدود الخطأ موجبا أم لا. وقد استخدم طريقة المربعات الدنيا العادية لتوفيق خط انحدار خطي للبيانات في الجدول (١٣-٢) والنتائج مبينة في نهاية الجدول (١٣-٢). وقد حصل عندئذ على الرواسب e_t، وهي مبينة في العمود الثالث من الجدول (١٣-٢). ومرسومة مقابل الزمن في الشكل (١٣-٣). لاحظ كيف أن الرواسب تقع بصورة متسقة فوق القيم التوفيقية وتحتها و ذلك لفترات غير قصيرة.

ويقترح مثل هذا النمط وجود ارتباط ذاتي موجب بين حدود الخطأ وذلك عند استخدام دالة انحدار مناسبة. لتأكيد هذا التشخيص البياني رغب المحلل في استخدام اختبار دربن ـ واتسون للبديلين:

$$H_0: \rho = 0$$
$$H_a: \rho > 0$$

والحسابات المطلوبة لإحصاءة الاختبار موجودة في الأعمدة (٤)، (٥) و(٦) من الجدول (١٣-٢). وبعد ذلك حصل المحلل على مايلي:

$$D = \frac{\sum_{t=2}^{20}(e_t - e_{t-1})^2}{\sum_{t=1}^{20} e_t^2} = \frac{.09794}{.13330} = .735$$

مستخدما مستوى معنوية 0.01 ولقيم $n = 20$ و $p - 1 = 1$ وجد من الجدول (أ-٦) ما يلي :

$$d_L = 0.95$$
$$d_U = 1.15$$

وبما أن قيمة $D = 0.735$ تقع تحت $d_L = 0.95$ فإن قاعدة القرار (13.14) تشير إلى أن

الاستنتاج المناسب هو H_a أي أن حدود الخطأ مرتبطة ذاتيا بصورة موجبة.

جدول (١٣-٢) بيانات مثال شركة بليسدل، نتائج الانحدار، وحسابات اختبار دربن - واتسون (بيانات مبيعات الشركة والصناعة معدلة فصليا)

(٦)	(٥)	(٤)	(٣) الراسب	(٢) مبيعات الصناعة (بملايين الدولارات)	(١) مبيعات الشركة (بملايين الدولارات)		السنة والفصل
e_t^2	$(e_t - e_{t-1})^2$	$e_t - e_{t-1}$	e_t	X_t	Y_t	t	
.0006787	-	-	-.026052	127.3	20.96	1	1983 :1
.0038459	.0012933	-.035963	-.062015	130.0	21.40	2	2
.0004849	.0070620	.084036	.022021	132.7	21.96	3	3
.0268154	.0200882	.141733	.163754	129.4	21.52	4	4
.0021688	.0137321	-.117184	.046570	135.0	22.39	5	1984 :1
.0021508	.0000000	-.000193	.046377	137.1	22.76	6	2
.0019024	.0000076	-.002760	.043617	141.2	23.48	7	3
.0034146	.0104146	-.102052	-.058435	142.8	23.66	8	4
.0089112	.0012934	-.035964	-.094399	145.5	24.10	9	1985 :1
.0222433	.0029968	-.054743	-.149142	145.3	24.01	10	2
.0219013	.0000013	.001151	-.147991	148.3	24.54	11	3
.0028147	.0090130	.094937	-.053054	146.4	24.30	12	4
.0005257	.0009076	.030126	-.022928	150.2	25.00	13	1986 :1
.0112046	.0165843	.128780	.105852	153.1	25.64	14	2
.0073041	.0004157	-.020388	.085464	157.3	26.36	15	[illegible]
.0112576	.0004259	.020638	.106102	160.7	26.98	16	[illegible]
.0008475	.0059275	-.076990	.029112	164.2	27.52	17	1987 :[illegible]
.0017906	.0001743	.013204	.042316	165.6	27.78	18	[illegible]
.0019501	.0074781	-.086476	-.044160	168.7	28.24	19	[illegible]
.0010896	.0001243	.011151	-.033009	171.7	28.78	20	[illegible]
.1333018	.0979400						المجموع

$$\hat{Y} = -1.4548 + .17628X$$

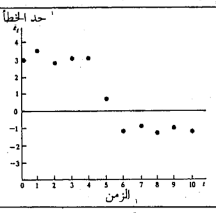
$$s\{b_0\} = .21415 \qquad s\{b_1\} = .00144$$

$$MSE = .00741$$

تعليقات

١- إذا احتجنا إلى اختبار للارتباط الذاتي السالب فإن إحصاءة الاختبار التي يمكن استخدامها هي $4 - D$، حيث D كما عرفناها من قبل. ويمكن إجراء الاختبار بالطريقة نفسها التي استخدمت لاختبار الارتباط الذاتي الموجب. أي أنه إذا كانت الكمية $4 - D$ واقعة تحت d_L فنستنتج أن $\rho < 0$ وبأنه يوجد ارتباط ذاتي سالب، وهكذا.

شكل (١٣-١) رسم الرواسب مقابل الزمن - مثال شركة بليسدل.



٢- يمكن القيام باختبار ثنائي الجانب للفرضية $H_0\colon \rho = 0$ ضد الفرضية $H_a : \rho \neq 0$ وذلك بالاستفادة من الاختبارين وحيدي الجانب كل على حدة. والخطأ من النوع الأول لاختبار ثنائي الجانب هو 2α، حيث α هو الخطأ من النوع الأول لكل من الاختبارين وحيدي الجانب.

٣- عندما يعطي اختبار دربن - واتسون الذي يستخدم الحدين d_L و d_U نتائج غير محددة، فإننا نحتاج من حيث المبدأ إلى مزيد من المشاهدات. وفي حالة بيانات سلاسل زمنية قد يكون من المستحيل، بالطبع، الحصول على مزيد من المشاهدات أو من الممكن أن تتوافر المشاهدات المطلوبة في المستقبل مما يؤدي إلى تأخير كبير عند الانتظار للحصول عليها. وعندما يكون اختبار الحدود غير مجدٍ فإن دربن- واتسون (مرجع [13.2]) يعطيان اختبارا تقريبيا يمكن استخدامه في هذه الحالة ولكن حتى يعطينا هذا الاختبار أكثر من مجرد مؤشر تقريبي لكون الارتباط الذاتي موجودا أم لا

ينبغي أن تكون درجات الحرية أكبر من 40.

وهناك طريقة معقولة تقضي باعتبار النتـائج غـير المحددة وكأنهـا تقـترح وجـود ارتباط ذاتي ومن ثم استخدام إحدى التدابير العلاجية، والتي سنناقشها فيما بعـد. وإذا كان هذا التدبير لايقودنا إلى نتائج انحدار مختلفـة اختلافـا كبـيرا فعندئـذ يمكـن الأخـذ بفرضية عدم وجود حدود خطأ مرتبطة. وعندمـا يـؤدي التدبـير العلاجـي بـالفعل إلى نتائج انحدار مختلفة تماما (كأن تعطى أخطاء معياريةمقدَّرة أكـبر لمعـاملات الانحـدار أو حذف الأخطاء المرتبطـة ذاتيـا)، فقـد تكـون النتـائج الـتي حصلنـا عليهـا مـن التدابـير العلاجية هي النتائج المناسبة.

٤ـ اختبار دربن ـ واتسون ليس منيعا ضد الخطأ في تحديد النموذج. مثـلا يمكـن ألا يكتشف الاختبار وجود أخطاء مرتبطة ذاتيا وتتبع نمط الانحـدار الذاتـي مـن المرتبـة الثانية المعرّف في (13.10).

٥ـ ومع الاستخدام الواسع لاختبار دربن ـــ واتسـون، إلا أنـه تتوافـر اختبـارات أخرى للارتباط الذاتي. وأحد هذه الاختبارات البديلة هو اختبار ثيل وناجار الموجـود في المرجع [13.3].

(١٣ـ٤) تدابير علاجية للارتباط الذاتي

التدبيران العلاجيان الرئيسان عند وجود حـدود خطـأ مرتبطـة ذاتيـا همـا إمـا بإضافـة واحد أو أكثر من المتغيرات المستقلة إلى نموذج الانحدار أو باستخدام متغيرات محوَّلة.

إضافة متغيرات مستقلة

أحد الأسباب الرئيسة لوجود حدود أخطاء مرتبطة ذاتيا، وكما بيّنا سـابقا، هـو حذف واحد أو أكثر من المتغيرات المستقلة من النموذج والتي لها تأثيرات مرتبـة زمنيـا على المتغير التابع. وعند اكتشاف وجود حدود خطأ مرتبطة ذاتيا فإن التدبير العلاجـي الأول يجب أن يكون دائما البحث عن متغيرات مستقلة أساسية محذوفـة. وقـد ذكرنـا سابقا عند الحديث عن انحدار المبيعات السنوية لمنتج ما على السـعر السـنوي المتوسـط لهذا المنتج في فترة 30 سنة بأن المتغير المحذوف هو حجـم المجتمـع، وفي بعـض الأحيـان

يمكن أن يساعد استخدام متغير ذي اتجاه خطي بسيط، أو استخدام متغيرات مؤشرة للتأثيرات الفصلية، في حذف الارتباط الذاتي في حدود الخطأ أو التقليل منه.

استخدام المتغيرات المحولة

لاينبغي استخدام تدبير علاجي يستند إلى تحويل المتغيرات إلا عندما لايكون استخدام متغيرات مستقلة إضافية مفيدا في التخلص من مشكلة الارتباط الذاتي للأخطاء. ولقد تم تطوير العديد من التدابير العلاجية المعتمدة على تحويل المتغيرات. وسنتطرق بالشرح لثلاث من هذه الطرق. وسيقتصر شرحنا على الانحدار الخطي البسيط ولكن التعميم إلى الانحدار المتعدد سهل ومباشر.

وتعتمد كل من هذه الطرق الثلاث على خاصية مهمة لنموذج الانحدار (13.1) حيث يكون لحد الخطأ انحدار ذاتي من المرتبة الأولى. لنأخذ المتغير المستقل المحوَّل (ويعني المتغير المستقل بعد التحويل):

$$Y_t' = Y_t - \rho Y_{t-1}$$

وبالتعويض عن Y_t و Y_{t-1} في هذه العبارة وفقا لنموذج الانحدار (13.1) نحصل على:

$$Y_t' = (\beta_0 + \beta_1 X_t - \varepsilon_t) - \rho(\beta_0 + \beta_1 X_{t-1} + \varepsilon_{t-1})$$

$$= \beta_0(1 - \rho) + \beta_1(X_t - \rho X_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

ولكن $u_t = \varepsilon_t - \rho\varepsilon_{t-1}$ وفقا لـ (13.1) وبالتالي:

$$Y_t' = \beta_0(1 - \rho) + \beta_1(X_t - \rho X_{t-1}) + \mu_t \qquad (13.15)$$

حيث u_t هي حدود الاضطرابات المستقلة. وهكذا عندما نستخدم المتغير Y_t'، فإن نموذج الانحدار يحوي حدود خطأ مستقلة. وعلاوة على ذلك فإن النموذج (13.15) هو نموذج انحدار خطي بسيط بمتغير مستقل جديد $X_t' = X_t - \rho X_{t-1}$ ويمكن ملاحظة ذلك بإعادة كتابة (13.15) كما يلي:

$$Y_t' = \beta_0' + \beta_1' X_t' + u_t \qquad (13.16)$$

حيث:

$$Y_t' = Y_t - \rho Y_{t-1}$$
$$X_t' = X_t - \rho X_{t-1}$$
$$\beta_0' = \beta_0 (1-\rho)$$
$$\beta_1' = \beta_1$$

وبالتالي نحصل، باستخدام المتغيرات المحوّلة X_t' و Y_t'، على نموذج انحدار خطي بسيط بحدود خطأ مستقلة. وهذا يعني أن طرق المربعات الدنيا العادية تحتفظ في هذا النموذج بخواصها المثلى المعتادة.

ولكي نتمكن من استخدام النموذج (13.16) فسنحتاج بصورة عامة، إلى تقدير معلمة الارتباط الذاتي ρ وذلك لأن قيمتها عادة غير معروفة. والطرق الثلاث التي سيتم وصفها تختلف في كيفية القيام بذلك. وعلى أي حال فالنتائج التي نحصل عليها من هذه الطرق الثلاث غالبا ماتكون متشابهة تماما. وحالما نحصل على تقدير لـ ρ، سنرمز له بـ r، فإننا نحصل على المتغيرات Y_t' و X_t'، مستخدمين هذا التقدير لـ ρ:

$$Y_t' = Y_t - rY_{t-1} \tag{13.17a}$$

$$X_t' = Y_t - rX_{t-1} \tag{13.17b}$$

وعندئذ نقوم بتوفيق نموذج الانحدار (13.16) لهذه البيانات المحولة مما ينتج دالة الانحدار المقدّرة:

$$\hat{Y}' = b_0' + b_1' X' \tag{3.18}$$

وإذا كانت دالة الانحدار التوفيقية قد ألغت الارتباط الذاتي في حدود الخطأ فإنه يمكننا العودة إلى نموذج انحدار توفيقي في المتغيرات الأصلية كما يلي:

$$\hat{Y} = b_0 + b_1 X \tag{13.19}$$

حيث:

$$b_0 = \frac{b_0'}{1-r} \tag{13.19a}$$

$$b_1 = b_1' \tag{13.19b}$$

ويمكن الحصول على الانحرافات المعيارية المقدرة لمعاملات الانحدار للمتغيرات الأصلية من مثيلاتها لمعاملات الانحدار للمتغيرات المحوّلة وذلك كما يلي:

$$s\{b_0\} = \frac{s\{b_0'\}}{1-r} \tag{13.20a}$$

$$s\{b_1\} = s\{b_1'\} \tag{13.20b}$$

طريقة كوكران ـ أوركت (*Cochran - Orcult*)

طريقة كوكران ـ أوركت هي طريقة تكرارية من ثلاث خطوات:

١ـ تقدير ρ: يمكن إنجاز هذه الخطوة بملاحظة أنه يمكن النظر إلى عملية الخطأ ذاتي الارتباط، المفترضة في النموذج (13.1) على أنها انحدار عبر نقطة الأصل:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

حيث ε_t هو المتغير التابع، ε_{t-1} المتغير المستقل، u_t حد الخطأ، وρ ميل الخط عبر نقطة الأصل. وبما أن ε_t و ε_{t-1} غير معروفة. فنستخدم الرواسب e_t و e_{t-1} التي حصلنا عليها بطرق المربعات الدنيا العادية، كمتغير تابع ومتغير مستقل على الترتيب، ونحصل على تقدير لـ ρ بتوفيق خط مستقيم عبر نقطة الأصل. ومن مناقشتنا السابقة للانحدار عبر نقطة الأصل نعلم بواسطة (13.15) أن تقدير الميل ρ ونرمز له بـ r يعطى بـ:

$$r = \frac{\sum_{t=2}^{n} e_{t-1}e_t}{\sum_{t=2}^{n} e_{t-1}^2} \tag{13.21}$$

٢ـ توفيق النموذج المحوّل (13.16): باستخدام قيمة ρ المقدَّرة من (13.21)، نحصل بعد ذلك على المتغيرات المحولة X_t' و Y_t' من (13.17). ونستخدم طريقة المربعات الدنيا العادية مع هذه المتغيرات المحولة لنحصل على دالة الانحدار التوفيقية (13.18).

٣ـ اختبار لمعرفة الحاجة للتكرار: نستخدم بعد ذلك اختبار دربن ـ واتسون لمعرفة ما إذا كانت حدود الخطأ للنموذج المحوّل غير مرتبطة. فإذا دل الاختبار على أنها غير مرتبطة تنتهي العملية. ونحصل على نموذج الانحدار التوفيقي بدلالة المتغيرات الأصلية عن طريق تحويل معاملات الانحدار مرة أخرى طبقا لـ (13.19).

إذا أشار اختبار دربن ـ واتسون إلى أن الارتباط الذاتي لايزال موجودا بعد التكرار الأول فعندئذٍ نعيد تقدير المعلمة ρ من الرواسب الجديدة لنموذج الانحدار

التوفيقي (13.19) بالمتغيرات الأصلية والتي حصلنا عليها من نموذج الانحدار التوفيقي (13.18) بالمتغيرات المحولة. ونحصل على مجموعة جديدة من المتغيرات المحولة بقيمة جديدة لـ r. ويمكن الاستمرار في هذه الطريقة لتكرار آخر أو تكرارين حتى يدل اختبار دربن ـ واتسون إلى أن حدود الخطأ في النموذج المحول غير مرتبطة. أما إذا لم تنته العملية بعد تكرار أو تكرارين فيجب عندئذ استخدام طريقة مختلفة.

مثال. في مثال شركة بليسدل، تظهر في الجدول (١٣-٣) الحسابات المطلوبة لتقدير معلمة الارتباط الذاتي ρ، والتي حصلنا عليها من الرواسب المحسوبة بطريقة المربعات الدنيا العادية مطبقة على المتغيرات الأصلية. ويعيد العمود الأول عرض الرواسب في الجدول (١٣-٢). ويحوي العمود الثاني الرواسب e_{t-1} أما الحسابات المطلوبة ففي العمودين الثالث والرابع. وبالتالي نقدّر r كما يلي:

$$r=\frac{.0834478}{.1322122}=.631166$$

ويمكننا الآن الحصول على المتغيرات المحولة Y_t' و X_t' في (13.17a) و(13.17b)

$$Y_t'=Y_t \ -.631166Y_{t-1}$$
$$X_t'=X_t-.631166X_{t-1}$$

ونتائج هذه الحسابات معروضة في الجدول (١٣-٤). حيث يوجد في العمودين الأول والثاني إعادة لقيم المتغيرات الأصلية Y_t و X_t فيما يحوي العمودان الثالث والرابع المتغيرات المحوَّلة Y_t' و X_t'. ونستخدم الآن توفيق انحدار خطي بطريقة المربعات الدنيا العادية لهذه المتغيرات المحولة مبنية على المشاهدات الـ $n-1$ المتبقية بعد التحويلات. وخط الانحدار التوفيقي مع نتائج الانحدار الأخرى معطاة في أسفل الجدول (١٣-٤). وخط الانحدار التوفيقي للمتغيرات المحوَّلة هو:

$$\hat{Y}'=-.3941+.17376X' \qquad (13.22)$$

حيث:

$$Y_t'=Y_t-.631166Y_{t-1} \qquad\qquad X_t'=X_t-.631166X_{t-1}$$

وبما أن الحد العشوائي من نموذج الانحدار المحول (13.16) هو حد الاضطراب u_t فإن $MSE = 0.00451$ هو تقدير لتباين حد الاضطراب هذا، تذكر أن $\sigma^2 = \sigma^2\{u_t\}$.

بناء على دالة الانحدار التوفيقية للمتغيرات المحولة (13.22) حصلنا على الرواسب ومن ثمَّ تمكنّا من حساب إحصاءة دربن - واتسون. النتيجة هي $D = 1.65$ (الحسابات غير مبينة هنا). ومن الجدول (أ-٦) وجدنا من أجل $\alpha = 0.01$، $p\text{-}1 = 1$ و$n = 19$.

$$d_L = 0.93 \qquad d_U = 1.13$$

وبما أن $D = 1.65 > d_U = 1.13$ فنستنتج أن معامل الارتباط لحدود الخطأ في النموذج ذي المتغيرات المحولة يساوي الصفر.

جدول (٣-١٣) الحسابات اللازمة لتقدير ρ بطريقة كوكران - أوركت - مثال شركة بليسدل.

t	(١) e_t	(٢) e_{t-1}	(٣) $e_{t-1}e_t$	(٤) e_{t-1}^2
1	-.026052	-	-	-
2	-.062015	-.026052	.0016156	.0006787
3	.022021	-.062015	-.0013656	.0038459
4	.163754	.022021	.0036060	.0004849
5	.046570	.163754	.0076260	.0268154
6	.046377	.046570	.0021598	.0021688
7	.043617	.046377	.0020228	.0021508
8	-.058435	.043617	-.0025488	.0019024
9	-.094399	-.058435	.0055162	.0034146
10	-.149142	-.094399	.0140789	.0089112
11	-.147991	-.149142	.0220718	.0222433
12	-.053054	-.147991	.0078515	.0219013
13	-.022928	-.053054	.0012164	.0028147
14	.105852	-.022928	-.0024270	.0005257
15	.085464	.105852	.0090465	.0112046
16	.106102	.085464	.0090679	.0073041
17	.029112	.106102	.0030889	.0112576
18	.042316	.029112	.0012319	.0008475
19	-.044160	.042316	-.0018687	.0017906
20	-.033009	-.044160	.0014577	.0019501
المجموع			.0834478	.1322122

$$r = \frac{\sum e_{t-1}e_t}{\sum e_{t-1}^2} = \frac{.0834478}{.1322122} = .631166$$

وبما أننا عالجنا بنجاح مشكلة حدود الخطأ المرتبطة ذاتيا فإننا نعود الآن فنحول

النموذج التوفيقي في (13.22) إلى المتغيرات الأصلية، مستخدمين (13.19):

$$b_0 = \frac{b_0'}{1-r} = \frac{-.3941}{1-.631166} = -1.0685$$

$$b_1 = b_1' = .17376$$

جدول (١٣-٤) المتغيرات المحولة ونتائج الانحدار للتكرار الأول وذلك بطريقة كوكران وأوركت - مثال شركة بليسدل.

(٤) $X_t' = X_t - .631166X_{t-1}$	(٣) $Y_t' = Y_t - .631166Y_{t-1}$	(٢) X_t	(١) Y_t	t
-	-	127.3	20.96	1
49.653	8.1708	130.0	21.40	2
50.648	8.4530	132.7	21.96	3
45.644	7.6596	129.4	21.52	4
53.327	8.8073	135.0	22.39	5
51.893	8.6282	137.1	22.76	6
54.667	9.1147	141.2	23.48	7
53.679	8.8402	142.8	23.66	8
55.369	9.1666	145.5	24.10	9
53.465	8.7989	145.3	24.01	10
56.592	9.3857	148.3	24.54	11
52.798	8.8112	146.4	24.30	12
57.797	9.6627	150.2	25.00	13
58.299	9.8608	153.1	25.64	14
60.668	10.1769	157.3	26.36	15
61.418	10.3425	160.7	26.98	16
62.772	10.4911	164.2	27.52	17
61.963	10.4103	165.6	27.78	18
64.179	10.7062	168.7	28.24	19
65.222	10.9559	171.7	28.78	20

$\hat{Y}' = -.3941 + .17376X'$

$s\{b_0'\} = .1672 \qquad s\{b_1'\} = .002957$

$MSE = .00451$

مما يقودنا إلى خط الانحدار التوفيقي في المتغيرات الأصلية.

$$\hat{Y} = -1.0685 + .17376X \qquad (13.23)$$

وأخيرا يمكننا الحصول على الانحرافات المعيارية المقدرة لمعاملات الانحدار للمتغيرات الأصلية باستخدام (13.20) وبناء على النتائج في الجدول (١٣-٤) نجد:

$$s\{b_0\}=\frac{s\{b_0'\}}{1-r}=\frac{.1672}{1-.631166}=.45332$$
$$s\{b_1\}=s\{b_1'\}=.002957$$

تعليقات

١- لا تعمل طريقة كوكران – أوركت، دائما بصورة مناسبة كما يجب. والسبب الرئيس لهذا هو أنه عندما تكون حدود الخطأ مرتبطة ذاتيا بصورة موجبة، تنحو قيمة r المقدَّرة في (13.21) إلى تقدير معلمة الارتباط الذاتي ρ بالنقصان.

وعندما يكون هذا الانحياز جدّيا فقد يقلل ذلك بشكل كبير من فعالية طريقة كوكران ـ أوركت.

٢- توجد علاقة تقريبية بين إحصاءة اختبار دربن ـ واتسون D في (13.13) وبين معلمة الارتباط المقدّرة r في (13.21):

$$D\approx 2(1-r) \tag{13.24}$$

وتشير هذه العلاقة إلى أن إحصاءة دربن ـ واتسون تتراوح تقريبا بين 0 و 4 وذلك لأن r تأخذ قيما بين 1- و1+، وأن قيمة D تساوي 2 عندما $r=0$. وفي مثال شركة بليسدل ومن توفيقة الانحدار بطريقة المربعات الدنيا وجدنا $D=0.735$، $r=0.631$ و $2(1-r)=0.738$.

٣- في ظروف معينة، قد يكون من المفيد وضع قيم شبه محولة للفترة الأولى، بحيث يستند الانحدار للمتغيرات المحولة على n بدلا من $n-1$ من المشاهدات وطرق القيام بذلك متوافرة في كتب متخصصة مثل المرجع [13.4].

٤- تنطبق خواص المربعات الدنيا للرواسب، مثل مجموع الرواسب يساوي الصفر، على رواسب دالة الانحدار التوفيقية للمتغيرات المحوَّلة، ولكن ليس لرواسب دالة الانحدار التوفيقية بعد إعادة تحويلها بدلالة المتغيرات الأصلية.

طريقة هيلد ريث ـ لُو

تتخذ طريقة هيلدريث ـ لُو لتقدير معلمة الارتباط الذاتي ρ، بغية استخدامها في التحويلات (13.17) الأسلوب نفسه الذي تتخذه طريقة بوكس ـ كوكس لتقدير

المعلمة λ في تحويل القوى لـ Y، بغية تحسين صلاحية نموذج الانحدار القياسي. إذ نختار في طريقة هيلدريث ـ لُو تلك القيمة لـ ρ التي تجعل مجموع مربعات الخطأ لنموذج الانحدار المحوّل (13.16) أصغر مايمكن:

$$SSE = \sum (Y_t' - \hat{Y}_t')^2 = \sum (Y_t' - b_0' - b_1' X_t')^2 \qquad (13.25)$$

وتتوافر برامج حاسب لايجاد قيمة ρ التي تجعل SSE أصغر مايمكن. وبصورة بديلة، يمكننا أن نبحث حسابيا بتشغيل انحدارات متكررة، معَ قيم مختلفة لـ ρ في كل انحدار، وذلك لاستطلاع القيمة التقريبية لـ ρ التي تجعل SSE أصغر مايمكن. وعند معرفة الفترة التي تقع فيها قيمة ρ التي تجعل SSE أصغر ما يمكن، يمكن البحث ضمن هذه الفترة على قيمة أكثر دقة لـ ρ.

وحالما نحصل على قيمة ρ التي تجعل SSE أصغر مايمكن يمكننا تحديد دالة الانحدار التوفيقية المقابلة لتلك القيمة لـ ρ. وإذا حُذفت عملية التحويل الارتباطَ الذاتي بنجاح فيمكن الحصول على دالة الانحدار التوفيقية في المتغيرات الأصلية بواسطة (13.19).

مثال. يحوي الجدول (١٣ـ٥) نتائج الانحدار لطريقة هيلدريث ــ لُو عند القيام بتوفيق نموذج الانحدار المحول (13.16) لبيانات شركة بليسدل وذلك من أجل قيم مختلفة لمعلمة الارتباط الذاتي. ونلاحظ أن SSE يأخذ أصغر قيمة له عندما تكون ρ قريبة من 0.96.

وهكذا سنأخذ $r = 0.96$ كتقدير لـ ρ. ودالة الانحدار التوفيقية للمتغيرات المحولة من أجل $r= 0.96$ بالإضافة إلى نتائج أخرى للانحدار، معطاة في أسفل الجدول (١٣ـ٥). ودالة الانحدار التوفيقية في المتغيرات المحوَّلة هي:

$$\hat{Y}' = .07117 + .16045X' \qquad (13.26)$$

حيث:

$$Y_t' = Y_t - .96Y_{t-1}$$
$$X_t' = X_t - .96X_{t-1}$$

وإحصاءة اختبار دربن ـ واتسون لهذا النموذج التوفيقي هي $D = 1.73$ ومن أجل القيم $n = 19$، $p - 1 = 1$ و $\alpha = 0.01$ تكون القيمة الحرجة العليا $d_U = 1.13$

ونستنتج أنه لم يبق ارتباط ذاتي في النموذج المحول.

ولذلك سنحول دالة الانحدار (13.26) عائدين مرة أخرى إلى المتغيرات الأصلية وباستخدام (13.19)، نحصل على:

$$\hat{Y} = 1.7793 + .16045X \qquad (13.27)$$

أما الانحرافات المعيارية المقدرة لمعاملات الانحدار فهي:

$$s\{b_0\} = 1.450 \qquad s\{b_1\} = .006840$$

جدول (١٣-٥) نتائج طريقة هيلدريث ـ لُو لمثال شركة بليسدل

SSE	ρ
.1170	.10
.0938	.30
.0805	.50
.0758	.70
.0728	.90
.0723	.92
.0718	.94
.07171	.95
.07167	.96
.07175	.97
.07197	.98

من أجل: $\rho = .96 \qquad \hat{Y}' = .07117 + .16045X'$

$$s\{b_0'\} = .05798 \qquad s\{b_1'\} = .006840$$

$$MSE = .00422$$

تعليقات

١ـ على العكس من طريقة كوكران ـ أوركت. لاتتطلب طريقة هيلدريث ـ لُو أي تكرارات عندما نحصل على تقدير لمعلمة الارتباط الذاتي ρ.

٢ـ لاحظ من الجدول (١٣ــ٥) أن SSE، كدالة في ρ، تبقى مستقرة تماما في منطقة واسعة حول القيمة الصغرى لها. وهذا هو مايحدث في الغالب، مما يشير إلى أنه لاحاجة لأن تكون عملية البحث الحسابية عن أفضل قيمة لـ ρ دقيقة جدا مالم يكن لدينا اهتمام خاص في حد التقاطع β_0، ذلك لأن التقدير b_0 حساس بالنسبة لقيمة r.

طريقة الفروق الأولى

بما أن قيمة معلمة الارتباط الذاتي ρ هي قيمة كبيرة في الغالب، وأن SSE كدالة في ρ تكون مستقرة تماما من أجل قيم كبيرة لـ ρ تصل حتى الـ 1.0 ، كما في مثال شركة بليسدل، فقد اقترح بعض الاقتصاديين والإحصائيين استخدام $\rho = 1.0$ في النموذج المحول (13.16) وإذا كانت $\rho = 1$ و $\beta'_0 = \beta_0(1-\rho) = 0$ فيصبح النموذج المحول (13.16) كما يلي:

$$Y'_t = \beta'_1 X'_t + u_t \qquad (13.28)$$

حيث:

$$Y'_t = Y_t - Y_{t-1} \qquad (13.28a)$$

$$X'_t = X_t - X_{t-1} \qquad (23.28b)$$

وهكذا نرى مرة أخرى أنه يمكن تقدير معاملات الانحدار مباشرة بطرق المربعات الدنيا العادية وترتكز هذه المرة على انحدار عبر نقطة الأصل. ونلاحظ أن المتغيرات المحولة في (13.2a) و(13.2b) هي فروقات أولى عادية. وقد وجد أن طريقة الفروقات الأولى فعّالة في تخفيض الارتباطات الذاتية لحدود الخطأ، وذلك في العديد من التطبيقات، وهي بالطبع أبسط بكثير من طريقتي كوكران ـ أوركت وهيلدريث ـ لُو.

ودالة الانحدار التوفيقية في المتغيرات المحولة هي:

$$\hat{Y}' = b'_1 X' \qquad (13.29)$$

ويمكن تحويلها والعودة مرة أخرى إلى المتغيرات الأصلية كما يلي:

$$\hat{Y} = b_0 + b_1 X \qquad (13.30)$$

حيث:

$$b_0 = \overline{Y} - b'_1 \overline{X} \qquad (13.20a)$$

$$b_1 = b'_1 \qquad (13.30b)$$

مثال. يحوي الجدول (١٣-٦) على المتغيرات المحوّلة و X بناء على تحويلات الفروقات الأولى في (13.28a,b) لمثال شركة بليسدل. ويؤدي تطبيق طريقة المربعات الدنيا لتقدير

الانحدار الخطي عبر نقطة الأصل إلى النتائج الموضحة في أسفل الجدول (١٣ــ٦). وخـط الانحدار التوفيقي في المتغيرات المحولة هو:

$$\hat{Y}' = .16849X' \qquad (13.31)$$

حيث:

$$Y_t' = Y_t - Y_{t-1}$$
$$X_t' = X_t - X_{t-1}$$

جدول (١٣-٦) الفروقات الأولى ونتائج الانحدار بطريقة الفروقات الأولى - مثال شركة بليسدل

(٤) $X_t' = X_t - X_{t-1}$	(٣) $Y_t' = Y_t - Y_{t-1}$	(٢) X_t	(١) Y_t	t
-	-	127.3	20.96	1
2.7	.44	130.0	21.40	2
2.7	.56	132.7	21.96	3
-3.3	-.44	129.4	21.52	4
5.6	.87	135.0	22.39	5
2.1	.37	137.1	22.76	6
4.1	.72	141.2	23.48	7
1.6	.18	142.8	23.66	8
2.7	.44	145.5	24.10	9
-.2	-.09	145.3	24.01	10
3.0	.53	148.3	24.54	11
-1.9	-.24	146.4	24.30	12
3.8	.70	150.2	25.00	13
2.9	.64	153.1	25.64	14
4.2	.72	157.3	26.36	15
3.4	.62	160.7	26.98	16
3.5	.54	164.2	27.52	17
1.4	.26	165.6	27.78	18
3.1	.46	168.7	28.24	19
3.0	.54	171.7	28.78	20

$$\hat{Y}' = .16849X'$$

$$s\{b_1'\} = .005096 \qquad MSE = .00482$$

ولفحص ما إذا كانت طريقة الفروقات الأولى قد أزالــت الارتبـاط الذاتـي فإننـا نستخدم اختبار دربن - واتسون. هناك نقطتان يجب ملاحظتهما عند استخدام اختبـار

دربن ـ واتسون مع طريقة الفروق الأولى. ففي بعض الأحيان، يمكن لطريقة الفروقات الأولى أن تفرط في التصحيح، مما يؤدي إلى ارتباط ذاتي سالب في حدود الخطأ. ولذلك، فقد يكون من المناسب استخدام اختبار دربن ـ واتسون ذي الجانبين عند اختبار الارتباط الذاتي لبيانات الفروق الأولى. والنقطة الثانية هي أنه لايوجد حد تقاطع في نموذج الفروق الأولى (13.28) بينما يتطلب اختبار دربن ـ واتسون نموذج انحدار توفيقي محتويا على حد تقاطع. يمكننا إجراء اختبار مشروع للارتباط الذاتي، في نموذج لايحوي حد تقاطع، وذلك بتوفيق دالة انحدار، لهذا الغرض، محتوية على حد تقاطع. وبالطبع، فإن النموذج الذي يهمنا هنا هو النموذج الذي لايحوي حد تقاطع.

وفي مثال شركة بليسدل، نجد أن إحصاءة دربن وواتسون لنموذج انحدار الفروق الأولى التوفيقي مع حد تقاطع، هي $D = 1.75$ وتشير هذه إلى عدم وجود ارتباط ذاتي في حدود الخطأ سواء باستخدام اختبار ذي جانب واحد ($\alpha = 0.01$) أو اختبار ذي جانبين ($\alpha = 0.02$).

وبما أن طريقة الفروقات الأولى قد أزالت وبنجاح الارتباط الذاتي، فإننا نعود إلى نموذج توفيقي في المتغيرات الأصلية باستخدام (13.30):

$$\hat{Y} = -.30349 + .16849X \qquad (13.32)$$

حيث:

$$b_0 = 24.569 - .16849(147.62) = -.30349$$

ونعلم من الجدول (١٣ـ٦) أن الانحراف المعياري المقدَّر لِـ b_1 هو $s\{b_1\} = 0.005096$ باعتبار أن $b_1 = b_1'$.

جدول (١٣ـ٧) نتائج الانحدار الرئيسة لطرق التحويل الثلاث ـ مثال شركة بليسدل

الطريقة	b_1	$s\{b_1\}$	r	تقدير σ^2 (MSE)
أوركت ـ كوكران	.1738	.0030	.63	.0045
هيلورث ـ لو	.1605	.0068	.96	.0042
الفروق الأولى	.1685	.0051	1.0	.0048
المتغيرات الأصلية	.1763	.0014	-	-

مقارنة بين الطرق الثلاث

يبين جدول (١٣-٧) بعض نتائج الانحدار الرئيسة للطرق الثلاث، بالإضافة إلى انحدار توفيقي للمتغيرات الأصلية بطريقة المربعات الدنيا العادية. ويمكننا ملاحظة النقاط الأساسية التالية:

١- تقديرات β_1 جميعها قريبة تماما بعضها من بعض.

٢- الانحرافات المعيارية المقدّرة لِـ b_1 المستندة إلى طرق هيلدريث - لُو وتحويل الفروق الأولى قريبة جدا بعضها من بعض بينما تكون هذه القيم أصغر إلى حد ما في طريقة كوكران ـ أوركت. ويبقى الانحراف المعياري المقدّر استنادا إلى الانحدار بطريقة المربعات الدنيا العادية، الانحراف المعياري الأصغر. وفي الحقيقة، كان هذا متوقعا، فقد ذكرنا سابقا أن الانحرافات المعيارية المقدَّرة $s\{b_k\}$ المحسوبة وفقا لطريقة المربعات الدنيا العادية يمكن أن تؤدي إلى فرط تقدير بالنقصان للانحرافات المعيارية الحقيقية $\sigma\{b_k\}$ وذلك عند وجود ارتباط ذاتي موجب.

٣- تعطي طرق التحويل الثلاث جميعها في الأساس، التقدير نفسه لِـ σ^2 أي لتباين حدود الاضطراب u_t.

ولاتعمل طرق التحويل الثلاث، دائما بالجودة نفسها، كما اتفق أن كانت الحالة هنا في مثال شركة بليسدل. وقد لاتزيل طريقة كوكران ـ أوركت الارتباط الذاتي في مجرد تكرار أو تكرارين وفي هذه الحالة يمكن أن تكون طريقة هيلدريث ـ لُو أو طريقة الفروقات الأولى هي الطريقة المفضلة وعندما يكون لعدد من طرق التحويل الفعالية نفسها في التخلص من الارتباط الذاتي فإننا نختار إحداها بناء على اعتبارات السهولة في الحسابات.

ملاحظة

لمزيد من التفصيل في مناقشة طرق كوكران ـ أوركت، هيلدريث ـ لُو والفروق الأولى بالإضافة إلى طرق علاجية أخرى للأخطاء المرتبطة ذاتيا يمكن الرجوع إلى كتب متخصصة مثل المرجع [13.4].

(١٣-٥) التنبؤ في حالة وجود حدود خطأ ذاتية الارتباط

القيام بتنبؤات أحد الاستخدامات المهمة لنماذج انحدار الخطأ ذاتي الانحدار. ففي هذه النماذج، يمكننا الاستفادة من المعلومات عن حد الخطأ في آخر فترة n للقيام بتنبؤ عن الفترة $n + 1$. وسيعطينا هذا تنبؤا أكثر دقة، لأنه عندما تكون نماذج انحدار الخطأ ذاتي الانحدار مناسبة فإن حدود الخطأ في الفترات المتتالية تكون مرتبطة. وهكذا إذا كانت المبيعات في الفترة n أعلى من قيمتها المتوقعة وكانت حدود الخطأ مرتبطة إيجابا، فمن المحتمل أن المبيعات في الفترة $n + 1$ ستكون أيضا أعلى من قيمتها المتوقعة.

وسوف نشرح الأفكار الأساسية وراء تطور التنبؤات بأن نستخدم مرة أخرى نموذج الانحدار الخطي البسيط مع حدود الخطأ ذاتية الانحدار (13.1). والتعميم إلى نموذج الانحدار المتعدد (13.2) سهل ومباشر، وسنستعرض أولا التنبؤ عندما تكون إما طرق كوكران ـ أوركت أو هيلدريث ـ لُو قد استُخدمت لتقدير معالم الانحدار.

عندما نعبر عن نموذج الانحدار (13.1):

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

باستخدام بنية حدود الخطأ:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

نحصل على:

$$Y_t = \beta_0 + \beta_1 X_t + \rho\varepsilon_{t-1} + u_t$$

وللفترة $n + 1$ نحصل على:

$$Y_{n+1} = \beta_0 + \beta_1 X_{n+1} + \rho\varepsilon_n + u_{n+1} \qquad (13.33)$$

وهكذا نرى أن Y_{n+1} تتكون من ثلاث مركبات:

١ـ القيمة المتوقعة $\beta_0 \pm \beta_1 X_{n+1}$

٢ـ عدد من المضاعفات ρ من حد الخطأ السابق ε_n

٣ـ حد اضطراب عشوائي مستقل بتوقع $E\{u_{n+1}\} = 0$.

ويمكن وضع التنبؤ للفترة التالية $n + 1$ وسنرمز له بـ F_{n+1}، بالتعامل مع كل من المركبات الثلاث في (13.33):

١- مع X_{n+1} معطاة، نقدِّر القيمة المتوقعة $\beta_0 + \beta_1 X_{n+1}$ كالمعتاد من دالة الانحدار التوفيقية:

$$\hat{Y}_{n+1} = b_0 + b_1 X_{n+1}$$

حيث b_0 وb_1 هي معاملات الانحدار المقدَّرة للمتغيرات الأصلية والتي حصلنا عليها من b_0' و b_1' للمتغيرات المحولة وفقا لـ (13.19).

٢- نقدِّر ρ بالقيمة r، وكذلك ε_n تُقدَّر بالرواسب e_n:

$$e_n = Y_n - (b_0 + b_1 X_n) = Y_n - \hat{Y}_n$$

ولهذا نقدر $\rho\varepsilon_n$ بالقيمة re_n.

٣- لحد الاضطراب u_{n+1} قيمة متوقعة صفر وهو مستقل عن المعلومات السابقة، وبالتالي فإننا نستخدم هذه القيمة المتوقعة في عملية التنبؤ.

ولهذا يكون التنبؤ للفترة $n + 1$:

$$F_{n+1} = \hat{Y}_{n+1} + re_n \quad (13.34)$$

يمكن الحصول على $(1-\alpha)$ فترة تنبؤ تقريبية لِ $Y_{n+1(new)}$ وهي المشاهدة الجديدة للمتغير التابع، وذلك باستخدام فترة التنبؤ المعتادة لمشاهدة جديدة في (3.35) ولكن استنادا إلى البيانات المحولة ولذلك فإنه يتم استبدال Y_t' و X_t' كما هما معرفتان في (13.17a,b) بـ Y_t وX_t في المعادلة (3.37a) للتباين المقدَّر $s^2\{Y_{h(new)}\}$.

ولذلك فإن $(1 - \alpha)$ حدي تنبؤ لـ$Y_{n+1(new)}$ مع انحدار خطي بسيط هما:

$$F_{n+1} \pm t(1-\alpha/2; n-3)s\{Y_{n+1(new)}\} \quad (13.35)$$

حيث $s\{Y_{n+1(new)}\}$ المعرفة في (3.37a)، مبنية الآن على البيانات المحولة. لاحظ أننا استخدمنا هنا $n - 3$ درجة حرية لعامل الضرب وذلك لأن لدينا $n - 1$ من البيانات المحولة بالإضافة إلى خسارة درجتي حرية من جراء تقدير المعلمتين في دالة الانحدار الخطي البسيط.

وعند القيام بتنبؤات تستند إلى طريقة الفروق الأولى فإن التنبؤ في (13.34) لايزال قابلا للتطبيق ولكن بعد أخذ $r = 1$، ويحسب الانحراف المعياري المقدَّر

$s\{Y_{n+1(new)}\}$ الآن وفقا لـ (5.21) لمتغير مستقل واحد وباستخدام المتغيرات المحوّلة. وأخيرا، فإن درجات الحرية لعامل الضرب t في (13.35) ستكون $n-2$، ذلك لأنه لايتم تقدير سوى معلمة واحدة في نموذج الانحدار (13.28) الذي لايوجد فيه حد تقاطع.

مثال

في مثال شركة بليسدل أشار إسقاط قامت به رابطة التجارة إلى أن المبيعات الصناعية في الربع الأول من العام ١٩٨٨ (أي الربع رقم 21) سيكون $X_{21} = 175.3$ مليونا من الدولارات.

وللتنبؤ بمبيعات شركة بليسدل في الربع رقم 21 سنستخدم دالة انحدار كوكران - أوركت التوفيقية (13.23):

$$\hat{Y} = -1.0685 + .17376X$$

في البداية ينبغي الحصول على الراسب e_{20}:

$$e_{20} = Y_{20} - \hat{Y}_{20} = 28.78 - [-1.0685 + .17376(171.7)] = .0139$$

والقيمة التوفيقية المقابلة لـ $X_{21} = 175.3$ هي:

$$\hat{Y}_{21} = -1.0685 + .17376(175.3) = 29.392$$

فيكون التنبؤ للفترة 21 عندئذ:

$$F_{21} = \hat{Y}_{21} + re_{20} = 29.392 + .631166(.0139) = 29.40$$

لاحظ كيف كان لحقيقة أن مبيعات الشركة في الربع 20 هي فوق معدلها بقليل، أثرها الإيجابي البسيط على التنبؤ لمبيعات الشركة في الربع 21:

ونرغب الآن في إقامة 95% فترة تنبؤ لـِ $Y_{21\,(new)}$ وباستخدام البيانات للمتغيرات المحولة في الجدول (١٣-٤)، نحسب $s\{Y_{n+1\,(new)}\}$ بواسطة (3.37) من أجل:

$$X'_{n+1} = X_n + 1 - .631166X_n = 175.3 - .631166(171.7) = 66.929$$

ونجد $s\{Y_{n+1(new)}\} = 0.0757$ (الحسابات غير موضحة هنا). ونحتاج لـ $t(.975;17) = 2.110$.

وبالتالي نحصل على حدي تنبؤ $29.40 \pm 2.110\,(.0757)$ وفترة تنبؤ:

$$29.24 \le Y_{21(new)} \le 29.56$$

ومع معرفة أن مبيعات الشركة المعدلة فصليا في الربع 20 كانت 28.78 مليونا

من الدولارات وأن المبيعات الصناعيــة في الربـع 21 كـانت 175.3 مليـون دولار فإننـا نتنبأ بمعامل ثقة تقريبي 95 في المئة أن مبيعات شركة بليسدل المعدلة فصليا في الربع 21 تقع بين 29.24 و29.56 مليون دولار.

وللحصول على تنبؤ للمبيعات الفعلية بما في ذلك التأثيرات الفصلية في الربع 21، فإنه ينبغي لشركة بليسدل أن تستوعب التأثـير الفصلـي لـلربع الأول في تنبـؤ المبيعـات المعدلة فصليا.

وتعطي طرق التحويل الأخرى تنبؤات مشابهة جدا للتنبؤات التي حصلنـا عليهـا من طريقة كوكران ـ أوركت. فمثلا باستخدام دالة الانحدار (13.32) المقـدرة بطريقـة الفروقات الأولى، يكون التنبؤ للربع 21 هو:

$$F_{21} = [-.30349 + .16849(175.3)] + 1.0[28.78 + .30349 - .16849(171.70)] = 29.39$$

والانحراف المعياري المقدَّر $s\{Y_{n+1(new)}\}$ محسوبا وفقا لـ (5.21) وبالاستناد إلى البيانات المحولة في الجدول (١٣-٦)، هو $s\{Y_{n+1(new)}\} = 0.0718$ (الحسـابات غـير موضحـة هنـا) وللحصول على 95 في المائة فترة تنبؤ، نحتاج لـ $t(0.975;18) = 2.101$ وبالتالي يكون حـدا التنبؤ: $29.39 \pm 2.101\ (.0718)$ وتكون الـ 95% فترة تنبؤ:

$$29.24 \le Y_{21(new)} \le 29.54$$

وهذا التنبؤ هو، عمليا، التنبؤ نفسه الذي حصلنا عليه من طريقة كوكران ـ أوركت.

والـ 95% فـترة تنبـؤ تقريبيـة باسـتخدام دالـة الانحـدار (13.27) المقـدَّرة بطريقـة هيلدريث ـ لُو (الحسابات غير موضحة هنا) هي:

$$29.24 \le Y_{21(new)} \le 29.52$$

وهذا التنبؤ هو عمليا، التنبؤ نفسه الذي حصلنا عليه من الطريقتين الأخريتين.

تعليقات

١ـ التنبؤات التي حصلنا عليها من نمـاذج الانحـدار (13.1) و(13.2) ذات حـدود الخطأ ذاتية الانحدار هي تنبؤات شرطية على المشاهدات السـابقة Y_n، Y_{n-1} وإلخ. وهـي كذلك شرطية على X_{n+1} والتي نضطر في الغالب إلى الحصول علـى قيمـة لهـا بطريقـة الإسقاط، وذلك كما في مثال شركة بليسدل.

٢- يمكن أيضا الحصول على تنبؤ لفترتين أو أكثر في المستقبل وذلك باستخدام علاقات ε_t الايقاعية بحدود الخطأ الماضية والتي طُوِّرت في الفقرة (١٣-٢). فمثلا، إذا أُعطينا X_{n+2} فإن التنبؤ للفترة $n + 2$، بناء على التقديرات من طريقة كوكران وأوركت أو طريقة هيلدريث - لُو، هو:

$$F_{n+2} = \hat{Y}_{n+2} + r^2 e_n \qquad (13.36)$$

وبالنسبة لطريقة الفروقات الأولى يُحسب التنبؤ في (13.36) مع أخذ $r = 1$.

٣- يفترض حدا التنبؤ التقريبيان (13.35) أن قيمة r المستخدمة في التحويلات (13.17a,b) هي القيمة الحقيقية لـ ρ، أي أن $r = \rho$. وإذا كانت الحال كذلك، فإن فرضيات الانحدار المعتادة لاتزال سارية وذلك لأننا نتعامل عندئذ مع النموذج المحول (13.16). ولرؤية أن حدي التنبؤ اللذين حصلنا عليهما من النموذج المحول لاتزال قابلة للتطبيق على التنبؤ F_{n+1} في (13.34) تذكر أن $\sigma^2\{Y_{h(new)}\}$ في (3.36) هو تباين الفرق $\hat{Y}_h - Y$ حيث Y هي المشاهدة الجديدة التي يُراد التنبؤ بقيمتها. ووفقا لمصطلحات حالتنا هنا مع المتغيرات المحولة نجد التقابل التالي:

$$\hat{Y}'_{n+1} = b'_0 + b'_1 X'_{n+1} = b_0(1-r)(X_{n+1} - rX_n) \quad \text{تقابل} \quad \hat{Y}_h ،$$

$$Y'_{n+1} = Y_{n+1} - rY_n \quad \text{تقابل} \quad Y ،$$

وبالتالي يكون الفرق $\hat{Y}'_{n+1} - Y'_{n+1}$ هو:

$$\begin{aligned}\hat{Y}'_{n+1} - Y'_{n+1} &= b_0(1-r) + b_1(X_n+1 - rX_n) - (Y_{n+1} - rY_n)\\ &= (b_0 + b_1X_n + 1) + r(Y_n - b_0 - b_1X_n) - Y_{n+1}\\ &= F_{n+1} - Y_{n+1}\end{aligned}$$

وهكذا تلعب F_{n+1} دور $\hat{Y}_h$ في (3.36)، وتلعب Y_{n+1} دور المشاهدة الجديدة Y. وحدا التنبؤ (13.35) هما حدان تقريبيان، ذلك لأن r ليست إلا تقديرا لـ ρ فقط.

المراجع

[13.1] Box, G. E. P. and Jenkins, G. M. *Time Series Analysis, Forecasting and control,* Rev. ed. San Francisco : Holden-Day, 1976.

[13.2] Durbin, J. and Watson, G. S. "Testing for Serial Correlation in Least Squares Regression. II."*Biometrika* 38 (1951), 159 - 78.

[13.3] Theil, H. and Nagar, A. L. "Testing the Independence of Regression

Disturbances." *Journal of the American Statistical Association* 56 (1961), 793 - 806.
[13.4] Pindyck, R. S. and Rubinfeld, D. L. *Econometric Models and Economic Forecasts*, 2nd ed. New York : McGraw-Hill, 1981.

مسائل

(١٣-١) بالرجوع إلى الجدول (١٣-١).

أ - ارسم ε_t مقابل ε_{t-1} لقيم $t = 1, 2, .., 10$ كيف يوضح الشكل الارتباط الذاتي الموجب من المرتبة الأولى في حدود الخطأ ؟.

ب - لو أنك رسمت u_t مقابل ε_{t-1} لقيم $t = 1, 2, 3, ..., 10$ فكيف تتوقع أن يكون الشكل؟

(١٣-٢) بالرجوع إلى **مسألة صلابة البلاستيك**، (٢-٢٠). لو أنه تم قياس وحدة الاختبار نفسها في 12 نقطة زمن مختلفة، فهل تتوقع أن تكون حدود الخطأ في نموذج الانحدار مرتبطة ذاتيا؟ ناقش.

(١٣-٣) صرح أحد الطلبة بقوله إن النماذج (13.1) و(13.2) ذات الخطأ ذاتي الانحدار من المرتبة الأولى غير ملائمة لبيانات السلاسل الزمنية في التجارة، لأن حدود الخطأ في الفترة t لهذه البيانات تخضع أيضا لتأثيرات عشوائية وقعت في أكثر من فترة واحدة من الماضي. علّق.

(١٣-٤) في أحد الاختبارات الفصلية استخدمت طالبة طريقة المربعات الدنيا لتوفيق نموذج انحدار خطي بسيط لبيانات سلاسل زمنية تحتوي على أخطاء مرتبطة ذاتيا بصورة موجبة وقد وجدت أن 90% فترة ثقة لِ β_1 كانت غير مفيدة لكونها واسعة جدا. وعندئذ قررت استخدام نموذج الانحدار (13.1) لزيادة دقة التقدير. علّق.

(١٣-٥) لكل من الاختبارات التالية والتي تتعلق بمعلمة الارتباط الذاتي ρ في نموذج الانحدار (13.2) بثلاثة متغيرات مستقلة، اذكر قاعدة القرار المناسبة بناء على إحصاءة دربن - واتسون لعينة حجمها 38:

$H_0: \rho = 0, H_a: \rho \neq 0, \alpha = .02;$ (١)

(٢) $H_0: \rho = 0,\ H_a: \rho < 0,\ \alpha = .05;$

(٣) $H_0: \rho = 0,\ H_a: \rho > 0,\ \alpha = .01;$

(١٣-٦) بالعودة إلى مسألة **صيانة الآلة الحاسبة**، (٢-١٨). حيث البيانات مرتبة زمنيا. افترض أن نموذج الانحدار (13.1) ملائم.اختبر وجود إرتباط ذاتي موجب من عدمه، استخدم$0.01 = \alpha$. اذكر الفرضيات البديلة، قاعدة القرار، والنتيجة.

(١٣-٧) بالرجوع إلى مسألة **شحنة الكيماويات**، (٧-١٢). البيانات مرتبة زمنيا. افترض أن نموذج الانحدار (13.2) ملائم. اختبر وجود ارتباط ذاتي موجب من عدمه، استخدم $\alpha = 0.05$ اذكر الفرضيات البديلة، قاعدة القرار، والنتيجة.

(١٣-٨) بالرجوع إلى مسألة **ناتج المحصول**، (٩-١٢). البيانات مرتبة زمنيا. افترض أن نموذج الانحدار (13.2) بحدود من المرتبة الأولى والثانية للمتغيرين العشوائيين المستقلين وعدم وجود حد تفاعل هو النموذج الملائم. اختبر وجود ارتباط ذاتي موجب من عدمه، استخدم $\alpha = 0.01$ اذكر الفرضيات البديلة، قاعدة القرار، والنتيجة.

(١٣-٩) **مركبات الحاسب الآلي**. أجرى أحد المحللين في مصنع لمركبات الحاسوب الصغيرة حصرا للبيانات الشهرية للستة عشر شهرا الماضية وذلك لقيمة الإنتاج الصناعي لوحدات المعالجة التي تستخدم هذه المركبات (X،بملايين الدولارات)، وكذلك لقيم المركبات المستخدمة من إنتاج المصنع (Y بآلاف الدولارات).

ويعتقد المحلل بأن علاقة انحدار خطي بسيط ستكون مناسبة هنا، ولكنه يتوقع وجود ارتباط ذاتي موجب. البيانات هي كالتالي:

أ - قم بتوفيق نموذج انحدار خطي بسيط بطريقة المربعات الدنيا العادية وأوجد الرواسب، وكذلك احسب $s\{b_0\}$ و$s\{b_1\}$.

ب - ارسم الرواسب مقابل الزمن ووضح ما إذا كنت تجد دليلا على وجود

ارتباط ذاتي موجب.

t	1	2	3	4	5	6	7	8
X_t	2.052	2.026	2.002	1.949	1.942	1.887	1.986	2.053
Y_t	102.9	101.5	100.8	98.0	97.3	93.5	97.5	102.2

t	9	10	11	12	13	14	15	16
X_t	2.102	2.113	2.058	2.060	2.035	2.080	2.102	2.150
Y_t	105.0	107.2	105.1	103.9	103.0	104.8	105.0	107.2

جـ -قم بإجراء اختبار للارتباط الذاتي الموجب مستخدما مستوى معنوية 0.05، اعرض الفرضيات البديلة، قاعدة القرار والنتيجة. هل يتفق تحليل الرواسب الذي أجريته في الفقرة (ب) مع نتيجة الاختبار؟

(١٣-١٠) بالرجوع إلى مسألة **مركبات الحاسوب**، (١٣-٩). قرر المحلل استخدام نموذج الانحدار (13.1) واستخدام طريقة كوكران – أوركت لتوفيق النموذج.

أ - احصل على تقدير نقطي لمعلمة الارتباط الذاتي. ماهي درجة صحة العلاقة التقريبية (13.24) بين هذا التقدير النقطي وإحصاءة اختبار دربن - واتسون؟

ب - استخدم تكرارا واحدا للحصول على التقديرات b_0' و b_1' لمعاملات الانحدار β_0' و β_1' في النموذج المحول (13.16) واعرض دالة الانحدار المقدرة. وكذلك احصل على $s\{b_0'\}$ و $s\{b_1'\}$.

جـ - اختبر ما إذا كان قد بقي أي ارتباط ذاتي موجب بعد التكرار الأول مستخدما مستوى معنوية 0.05. اذكر الفرضيات البديلة، قاعدة القرار والنتيجة.

د - أعد عرض دالة الانحدار المقدرة التي حصلت عليها في الجزء (ب) بدلالة المتغيرات الأصلية وكذلك أوجد $s\{b_0\}$ و $s\{b_1\}$. قارن معاملات الانحدار المقدرة وانحرافاتها المعيارية المقدرة التي حصلت عليها من طريقة كوكران - أوركت بتلك التي تم الحصول عليها من طريقة المربعات الدنيا العادية في المسألة (١٣-٩)أ.

هـ - بناء على النتائج من الجزئين (جـ) و(د) هل يبدو أن طريقة كوكران - أوركت كانت فعالة هنا؟.

و - ستكون قيمة الإنتاج الصناعي في الشهر 17 مبلغ 2.210 مليون دولار. تنبأ بقيمة المركبات من إنتاج المصنع التي استُخدمت في الشهر 17 استخدم 95% فترة تنبؤ، وفسر هذه الفترة.

ز - قدر β_1 بـ 95% فترة ثقة. فسّر هذه الفترة.

(١٣-١١) بالرجوع إلى مسألة **مركبات الحاسوب**، (١٣-٩). افترض أن نموذج الانحدار (13.1) قابل للتطبيق.

أ - استخدم طريقة هيلدريث - لُو للحصول على تقدير نقطي لمعلمة الارتباط الذاتي. قم ببحث حسابي مستخدما القيم $\rho = 0.1, 0.2, \ldots, 1.0$ اختر منها قيمة ρ التي تجعل SSE أصغر ما يمكن.

ب - بناء على التقدير الذي حصلت عليه من الجزء (أ)، احصل على تقدير لدالة الانحدار المحولة (13.16). وكذلك أوجد $s\{b_0'\}$ و $s\{b_1'\}$.

جـ - قم بإجراء اختبار لتقدير ما إذا كان الارتباط الذاتي الموجب لايزال باقيا في نموذج الانحدار المحول، استخدم $\alpha = .05$ اذكر الفرضيات البديلة، قاعدة القرار والنتيجة.

د - اعرض دالة الانحدار المقدرة التي حصلت عليها من الجزء (ب) بدلالة المتغيرات الأصلية. وكذلك أوجد $s\{b_o\}$ و $s\{b_1\}$. قارن معاملات الانحدار المقدرة وانحرافاتها المعيارية التي حصلت عليها من طريقة هيلدريث - لُو بتلك التي تم الحصول عليها من طريقة المربعات الدنيا العادية (13.9a).

هـ - بناء على النتائج من الجزئين (جـ) و (د) هل يبدو أن طريقة هيلدريث - لُو كانت فعّالة هنا؟.

و - ستكون قيمة الإنتاج الصناعي في الشهر 17 هي 2.210 مليون دولار تنبأ بقيمة المركبات من إنتاج المصنع التي استخدمت في الشهر 17. استخدم 95% فترة تنبؤ. فسّر هذه الفترة.

ز - قدر β_1 بـ 95% فترة ثقة. فسّر هذه الفترة.

(١٣-١٢) بالرجوع إلى مسألة **مركبات الحاسوب الآلي**. (١٣-٩). افترض أن نموذج الانحدار (13.1) قابل للتطبيق وأننا نرغب في استخدام طريقة الفروقات الأولى.

أ - قدّر معامل الانحدار β_1' في نموذج الانحدار المحول (13.28)، ومن ثَمَّ أوجد تقديرا للانحراف المعياري لهذا التقدير. اكتب دالة الانحدار المقدّرة.

ب - اختبر ما إذا كانت حدود الخطأ في طريقة الفروقات الأولى مرتبطة ذاتيا مستخدما اختبارا ذا جانبين ومستوى معنوية 0.10. اذكر الفرضيات البديلة، قاعدة القرار، والنتيجة. وضح لماذا يكون للاختبار ذي الجانبين دلالة هنا؟

جـ - أعد عرض دالة الانحدار المقدرة التي حصلت عليها في الجزء (أ) بدلالة المتغيرات الأصلية. وأوجد $s\{b_1\}$. قارن معاملات الانحدار المقدَّرة وانحرافاتها المعيارية التي حصلت عليها من طريقة الفروق الأولى بتلك التي تم الحصول عليها من طريقة المربعات الدنيا العادية. في المسألة (١٣-٩)أ.

د - بناء على النتائج من الجزئين (ب) و(جـ) هل يبدو أن طريقة الفروقات الأولى كانت فعّالة هنا؟

هـ - ستكون قيمة الإنتاج الصناعي في الشهر 17 هي 2.210 مليون دولار تنبأ بعدد المركبات من إنتاج المصنع التي استخدمت في الشهر 17. استخدم 95% فترة تنبؤ. فسّر هذه الفترة.

و - قدر β_1 بـ 95% فترة ثقة. فسر هذه الفترة.

(١٣-١٣) **وكالة إعلان**. يرغب الشريك الإداري لوكالة إعلانات في معرفة إمكانية

القيام بتنبؤات دقيقة للفواتير الشهرية. والبيانات الشهرية لقيمة الفواتير (Y بآلاف الدولارات الثابتة) ولعدد ساعات العمل (X بآلاف الساعات) للأشهر العشرين الأخيرة موضحة في الجدول التالي. ويُعتقد بأنه من الممكن أن يكون نموذج انحدار خطي بسيط مناسب هنا، ولكن من الممكن أن تكون حدود الخطأ مرتبطة ذاتيا بصورة موجبة.

7	6	5	4	3	2	1	t
3.020	2.523	2.305	2.524	2.234	2.171	2.521	X_t
247.6	222.7	211.3	221.9	207.2	203.9	220.4	Y_t
14	13	12	11	10	9	8	t
3.586	3.576	3.801	3.737	3.461	3.532	3.014	X_t
275.1	275.4	287.0	283.9	269.1	272.9	247.6	Y_t
	20	19	18	17	16	15	t
	3.618	3.623	3.117	3.019	2.723	3.447	X_t
	278.5	278.6	252.4	248.1	232.8	269.1	Y_t

أ ـ قم بتوفيق نموذج انحدار خطي بسيط بطريقة المربعات الدنيا العادية واحصل على الرواسب. كذلك أوجد $s\{b_0\}$ و $s\{b_1\}$.

ب ـ ارسم الرواسب مقابل الزمن ووضح ما إذا كنت تجد دليلا على وجود ارتباط ذاتي موجب.

جـ ـ قم بإجراء اختبار للارتباط الذاتي الموجب مستخدما مستوى معنوية 0.01. اذكر الفرضيات البديلة، قاعدة القرار والنتيجة. هل يتفق تحليل الرواسب الذي أجريته في الفقرة (ب) مع نتيجة الاختبار؟

(١٣-١٤) بالرجوع إلى مسألة **وكالة الإعلان** (١٣-١٣). افترض أن نموذج الانحدار (13.1) قابل للتطبيق وأننا نرغب في استخدام طريقة كوكران ـ أوركت.

أ ـ أوجد تقديرا نقطيا لمعلمة الارتباط الذاتي. ماهي درجة صحة العلاقة التقريبية (13.24) بين هذا التقدير النقطي وإحصاءة اختبار دربن ـ واتسون؟

ب ـ استخدم تكرارا واحدا للحصول على التقديرين b_0' و b_1' لمعلمتي الانحدار β_0' و β_1' في النموذج المحول (13.16) واكتب دالة الانحدار

المقدّرة. أوجد أيضا $s\{b_0'\}$ و $s\{b_1'\}$.

جـ - اختبر ما إذا كان أي ارتباط ذاتي موجب قد بقي بعد التكرار الأول مستخدما مستوى معنوية 0.01. اذكر الفرضيات البديلة، قاعدة القرار والنتيجة.

د - أعد كتابة دالة الانحدار المقدّرة والتي حصلت عليها من الجزء (ب) بدلالة المتغيرات الأصلية. كذلك احصل على $s\{b_0\}$ و $s\{b_1\}$. قارن معاملات الانحدار المقدّرة وانحرافاتها المعيارية والتي حصلت عليها من طريقة كوكران - أوركت بتلك التي تم الحصول عليها من طريقة المربعات الدنيا العادية في المسألة (١٣-١٣)أ.

هـ - بناء على النتائج في الجزئين (جـ) و(د) هل يبدو أن طريقة كوكران - أوركت كانت فعّالة هنا؟

و - يُتوقع في الشهر 21 أن يكون عدد ساعات العمل 3.625 ألف ساعة. قم بالتنبؤ بقيمة الفواتير بآلاف الدولارات الثابتة في الشهر 21 مستخدما 99% فترة تنبؤ - فسّر هذه الفترة.

ز - قدر β_1 بـ 99% فترة ثقة - فسّر هذه الفترة.

(١٣-١٥) بالرجوع إلى مسألة **وكالة الإعلان** (١٣-١٣). افترض أن نموذج الانحدار (13.1) مناسب للتطبيق.

أ - استخدم طريقة هيلدريث - لُو للحصول على تقدير نقطي لمعلمة الارتباط الذاتي. قم ببحث حسابي مستخدما القيم $\rho = 0.1, 0.2, \ldots, 1.0$ واختر منها القيمة التي تجعل SSE أصغر ما يمكن.

ب - بناء على التقدير الذي حصلت عليه من الجزء (أ)، احصل على تقدير لدالة الانحدار المحولة (13.16). أوجد أيضا $s\{b_0'\}$ و $s\{b_1'\}$.

جـ - قم بإجراء اختبار لتقرير ما إذا كان الارتباط الذاتي الموجب لايزال

باقيا في نموذج الانحدار المحول، استخدم $\alpha = 0.01$. اعرض الفرضيات البديلة، قاعدة القرار والنتيجة.

د - أعد كتابة دالة الانحدار المقدرة التي حصلت عليها من الجزء (ب) بدلالة المتغيرات الأصلية. وأوجد $s\{b_0\}$ و $s\{b_1\}$ قارن معاملات الانحدار المقدرة وانحرافاتها المعيارية المقدرة التي حصلت عليها من طريقة هيلدريث - لو بتلك التي تم الحصول عليها من طريقة المربعات الدنيا العادية في المسألة (١٣-١٣)؟.

هـ - بناء على النتائج في الجزئين (جـ) و(د) هل يبدو أن طريقة هيلدريث ولو كانت فعّالة هنا ؟

و - يتوقع في الشهر 21 أن يكون عدد ساعات العمل 3.625 ألف ساعة. تنبأ بقيم الفواتير بآلاف الدولارات الثابتة في الشهر 21 مستخدما 99 في المائة فترة تنبؤ. فسّر هذه الفترة.

ز - قدر β_1 بـ 99% فترة ثقة. فسّر هذه الفترة.

(١٣-١٦) بالرجوع إلى مسألة **وكالة الإعلان**، (١٣-١٣). افترض أن نموذج الانحدار (13.1) قابل للتطبيق وأننا نرغب في استخدام طريقة الفروق الأولى.

أ - قدّر معامل الانحدار في نموذج الانحدار β_1' المحول (13.28) وأوجد الانحراف المعياري المقدَّر لهذا التقدير. اكتب دالة الانحدار المقدَّرة.

ب - اختبر ما إذا كانت حدود الخطأ في طريقة الفروق الأولى مرتبطة ذاتيا وذلك باستخدام اختبار ذي جانبين ومستوى معنوية 0.02 اذكر الفرضيات البديلة، قاعدة القرار والنتيجة. وضح لماذا يكون للاختبار ذي الجانبين دلالة هنا؟

جـ - أعد كتابة دالة الانحدار المقدّرة التي حصلت عليها في الجزء (أ) بدلالة المتغيرات الأصلية. أوجد أيضا $s\{b_1\}$. قارن معاملات الانحدار المقدَّرة التي حصلت عليها من طريقة الفروق الأولى والانحراف

المعياري المقدّر بتلك التي تم الحصول عليها من طريقة المربعات الدنيا العادية في المسألة (١٣-١٣)أ.

د - بناء على النتائج من الجزئين (ب) و (جـ) هل يبدو أن طريقة الفروقات الأولى كانت فعّالة هنا ؟

هـ - يتوقع في الشهر 21 أن يكون عدد ساعات العمل 3.625 ألف ساعة. تنبأ بقيم الفواتير بآلاف الدولارات الثابتة في الشهر 21 مستخدما فترة 99%. فسّر هذه الفترة.

و - قدر β_1 بـ 99% فترة ثقة. فسّر هذه الفترة.

(١٣-١٧) مبيعات **شركة مكجيل**. تبين البيانات أدناه المبيعات ربع السنوية المعدلة فصليا لمبيعات شركة مكجيل (Y بملايين الدولارات) وكذلك لمبيعات الصناعة إجمالا (X بملايين الدولارات) وذلك لآخر عشرين ربعا الماضية.

7	6	5	4	3	2	1	t
141.1	137.1	135.0	129.4	132.7	130.0	127.3	X_t
23.48	22.76	22.39	21.52	21.96	21.40	20.96	Y_t

14	13	12	11	10	9	8	t
153.1	150.2	146.4	148.3	145.3	145.5	142.8	X_t
25.64	25.00	24.28	24.54	24.01	24.10	23.66	Y_t

20	19	18	17	16	15	t
172.0	168.7	165.6	164.2	160.7	157.3	X_t
28.78	28.24	27.78	27.52	26.98	26.46	Y_t

أ - هل تتوقع أن تكون معلمة الارتباط الذاتي ρ موجبة، سالبة، أم صفرا؟

ب - قم بتوفيق نموذج انحدار ذاتي بسيط بطريقة المربعات الدنيا العادية واحسب الرواسب. كذلك احصل على $s\{b_0\}$ و $s\{b_1\}$.

جـ - ارسم الرواسب مقابل الزمن ووضح ما إذا كنت تجد دليلا على وجود ارتباط ذاتي موجب.

د - قم بإجراء اختبار للارتباط الذاتي الموجب مستخدما مستوى معنوية $\alpha = 0.01$. اذكر الفرضيات البديلة، قاعدة القرار والنتيجة. هل يتفق

تحليل الرواسب الذي أجريته في الفقرة (جـ) مع نتيجة الاختبار؟.

(١٣-١٨) بالرجوع إلى مسألة **مبيعات شركة مكجيل**. (١٣-١٧). افترض أن نموذج الانحدار (13.1) قابل للتطبيق وأننا نرغب في استخدام طريقة كوكران ــ أوركت.

أ ـ أوجد تقديرا نقطيا لمعلمة الارتباط الذاتي. ماهي درجة صحة العلاقة التقريبية (13.24) بين هذا التقدير النقطي وإحصاءة اختبار دربن ــ واتسون؟

ب ـ استخدم تكرارا واحدا للحصول على التقديرات b_0' و b_1' لمعلمتي الانحدار β_0' و β_1' في نموذج الانحدار المحول (13.16) واكتب دالة الانحدار المقدَّرة. كذلك احسب $s\{b_0\}$ و $s\{b_1'\}$.

جـ ـ قم باختبار لوجود ارتباط ذاتي موجب بعد التكرار الأول مستخدما مستوى معنوية $\alpha = 0.01$. اذكر الفرضيات البديلة، قاعدة القرار والنتيجة.

د ـ أعد كتابة دالة الانحدار المقدَّرة التي حصلت عليها من الجزء (ب) بدلالة المتغيرات الأصلية. واحسب أيضا $s\{b_0\}$ و$s\{b_1\}$ قارن دالة الانحدار المقدّرة وانحرافاتها المعيارية المقدّرة التي حصلت عليها من طريقة كوكران وأوركت بتلك التي تم الحصول عليها من طريقة المربعات الدنيا العادية في المسألة (١٣-١٧)ب.

هـ ـ بناء على النتائج من الجزئين (جـ) و(د) هل يبدو أن طريقة كوكران– أوركت كانت فعّالة هنا؟.

و ـ يتوقع في الربع 21 أن تكون المبيعات الصناعية 181.0 مليون دولار. تنبأ بمبيعات شركة مكجيل في الربع 21 مستخدما 99% فترة تنبؤ. فسر هذه الفترة.

ز ـ قدر β_1 بـ 90% فترة ثقة. فسّر هذه الفترة.

(١٣-١٩) بالرجوع إلى مسألة **مبيعات شركة مكجيل** (١٣-١٧). افترض أن نموذج الانحدار (13.1) قابل للتطبيق.

أ - استخدم طريقة هيلدريث - لُو للحصول على تقدير نقطي لمعلمة الارتباط الذاتي. قم ببحث حسابي مستخدما القيم $\rho = 0.1, 0.2, ..., 1.0$ واختر منها قيمة ρ التي تجعل SSE أصغر مايمكن.

ب - بناء على التقدير الذي حصلت عليه من الجزء (أ) أوجد تقديرا لدالة الانحدار المحولة وكذلك أوجد $s\{b'_0\}$ و $s\{b'_1\}$.

جـ - قم باختبار لتقرير ما إذا كان الارتباط الذاتي الموجب لايزال باقيا في نموذج الانحدار المحول، مستخدما $\alpha = 0.01$. اذكر الفرضيات البديلة، قاعدة القرار والنتيجة.

د - أعد كتابة دالة الانحدار المقدّرة التي حصلت عليها من الجزء (ب) بدلالة المتغيرات الأصلية. وكذلك أوجد $s\{b_0\}$ و$s\{b_1\}$. قارن معاملات الانحدار المقدَّرة وانحرافاتها المعيارية المقدَّرة التي حصلت عليها من طريقة هيلدريث - لُو بتلك التي تم الحصول عليها من طريقة المربعات الدنيا العادية في المسألة (١٣-١٧)ب.

هـ - بناء على النتائج في الجزئين (جـ) و(د) هل يبدو أن طريقة هيلدريث - لُو كانت فعّالة هنا ؟

و - يتوقع في الربع 21 أن تكون المبيعات الصناعية 181.0 مليون دولار. تنبأ بمبيعات شركة مكجيل في الربع 21 مستخدما 90% فترة تنبؤ. فسّر هذه الفترة.

ز - قدر β_1 بـ 90% فترة ثقة. فسر هذه الفترة.

(١٣-٢٠) بالرجوع إلى مسألة **مبيعات شركة مكجيل** (١٣-١٧). افترض أن نموذج الانحدار (13.1) قابل للتطبيق، وأننا نرغب في استخدام طريقة الفروق الأولى.

أ - قدّر معامل الانحدار في نموذج الانحدار β'_1 المحول (13.28) واحسب

الانحراف المعياري المقدّر لهذا التقدير. اكتب دالة الانحدار المقدّرة.

ب ـ اختبر ما إذا كانت حدود الخطأ في طريقة الفروق الأولى مرتبطة ذاتيا بصورة موجبة وذلك باستخدام اختبار ذي جانبين ومستوى معنوية $\alpha = 0.01$. اعرض الفرضيات البديلة، قاعدة القرار، والنتيجة.

جـ ـ أعد كتابة دالة الانحدار المقدَّرة التي حصلت عليها من الجزء (أ) بدلالة المتغيرات الأصلية. وكذلك احسب $s\{b_1\}$. قارن معاملات الانحدار المقدَّرة والانحراف المعياري المقدَّر $s\{b_1\}$. التي حصلت عليها من طريقة الفروق الأولى بتلك التي تم الحصول عليها من طريقة المربعات العادية في المسألة (١٣-١٧)ب.

د ـ بناء على النتائج في الجزئين (ب) و(جـ) هل يبدو أن طريقة الفروق الأولى كانت فعّالة هنا؟.

هـ ـ يُتوقع في الربع 21 أن تكون المبيعات الصناعية 181.0 مليون دولار. تنبأ بمبيعات شركة مكجيل في الربع 21 مستخدما 90% فترة تنبؤ. فسّر هذه الفترة.

و ـ قدر β_1 بـ 90% فترة ثقة. فسّر هذه الفترة.

(١٣-٢١) قام أحد الطلاب بتطبيق لتحويلات الفروق الأولى في (13.28a,b) ووجد أن عدة قيم لـ X_t' تساوي الصفر ولكن القيم المقابلة لـ Y_t' لم تكن تساوي الصفر. فهل يعني هذا أن تحويلات الفروق الأولى غير صالحة للبيانات ؟.

تمارين

(١٣-٢٢) استنبط (13.8) في حالة $s = 2$.

(١٣-٢٣) بالرجوع إلى نموذج الخطأ ذاتي الانحدار من المرتبة الأولى (13.1). افترض أن Y_t هي النسبة المئوية لحصة الشركة في السوق، X_t سعر البيع للشركة كنسبة مئوية من متوسط سعر البيع المنافس، $\beta_0 = 100$ ، $\beta_1 = -0.35$ ، $\rho = 0.6$ ،

$\sigma^2 = 1$ و$\varepsilon_0 = 2.403$. وليكن X_t و u_t لقيم $t = 1, 2,..., 10$ كما يلي:

t	1	2	3	4	5	6	7	8	9	10
X_t	100	115	120	90	85	75	70	95	105	110
u_t	.764	.509	-.242	-1.808	-.485	.501	-.539	.434	-.299	.030

أ - ارسم خط الانحدار الحقيقي. ولّد المشاهدات $Y_t(t = 1,..,10)$ وارسم هذه القيم على الشكل نفسه. قم بتوفيق خط الانحدار لهذه البيانات بطريقة المربعات الدنيا وارسمه على الشكل نفسه. ماهي الصلة بين خط الانحدار التوفيقي والخط الحقيقي؟

ب - كرر الخطوات نفسها في (أ) ولكن هذه المرّة لتكن $\rho = 0$. في أي من هاتين الحالتين يكون خط الانحدار التوفيقي أقرب إلى الخط الحقيقي؟ هل هذه النتيجة التي حصلت عليها هي النتيجة المتوقعة؟

جـ - قم بتوليد البيانات Y_t من أجل $\rho = -0.7$ ولكل من القيم $\rho = 0.6$، $\rho = 0$ و$\rho = -0.7$. احصل على فروق حد الخطأ المتتابعة $\varepsilon_t - \varepsilon_{t-1}(t = 1,...,10)$.

د - في أي من الحالات الثلاث في الفقرة (جـ) يكون $\sum(\varepsilon_t - \varepsilon_{t-1})^2$ هو الأصغر؟ وفي أي منها يكون الأكبر ؟ كيف تقترح أن يكون التعميم في هذه الحالة؟

(١٣-٢٤) لنموذج الانحدار المتعدد (13.2) بـ $p - 1 = 2$، استنبط النموذج المحول الذي تكون فيه الحدود العشوائية غير مرتبطة.

(١٣-٢٥) افترض أن عملية الخطأ ذاتي الانحدار للنموذج $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$ هي تلك المعطاة في (13.10).

أ - كيف ستكون المتغيرات المحوّلة Y_t' و X_t' في نموذج الانحدار الذي تكون الحدود العشوائية فيه غير مرتبطة؟

ب - كيف ستقدر المعلمتين ρ_1 وρ_2 لاستخدامهما في طريقة كوكران وأوركت؟

جـ - كيف ستقدر المعلمتين ρ_1 وρ_2 لاستخدامهما في طريقة هيلدريث - لُو؟

(١٣-٢٦) استنبط التنبؤ F_{n+1} لنموذج الانحدار الخطي البسيط مع نمـوذج خطـأ ذاتـي الانحدار من المرتبة الثانية (13.10).

مشاريع

(١٣-٢٧) ليكن نموذج الانحدار الصحيح هو $Y_t = 10 + 24\,X_t + \varepsilon_t$ حيث:
$\varepsilon_t = 8\varepsilon_{t-1} + u_t$ و u_t مستقلة و $N(0.25)$.

أ - قم بتوليد 11 رقما عشـوائيا مسـتقلا مـن $N(0.25)$. اسـتخدم الرقـم العشوائي على أنه ε_0 واكتب حدود الخطأ العشرة $\varepsilon_1,\ldots,\varepsilon_{10}$ وعندئذ احسب قيم المشاهدات العشرة $Y_1,\ldots,Y_{10}$ التي تقابل القيـم $X_1 = 1,$ $X_2 = 2,\ldots, X_{10} = 10$ قم بتوفيق دالة انحدار خطي بطريقـة المربعـات الدنيا العادية واحسب MSE.

ب - أعد الجزء (أ) 100 مرة مستخدما أرقاما عشوائية جديدة في كل مرّة.

جـ - احسب متوسـط التقديـرات المائـة b_1. هـل يبـدو أن b_1 مقـدِّر غـير منحاز لِـ β_1 بالرغم من وجود ارتباط ذاتي موجب.

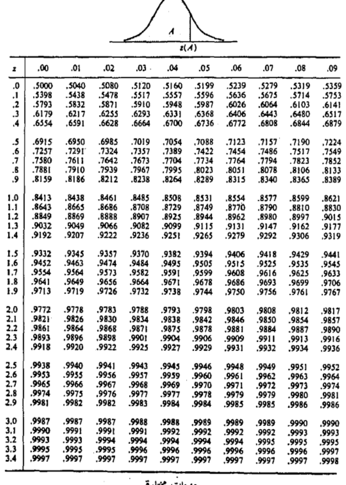
د - احسب متوسط التقديرات المائة لـ MSE. هل يبـدو أن MSE مقـدِّر منحاز لِـ σ^2؟ إذا كان الأمر كذلك، فهل يبدو مقدار الانحياز صغـيرا أم كبيرا؟.

الملحق ا

جداول

جدول (ا-١) الاحتمالات المتجمعة للتوزيع الطبيعي المعياري.

العدد في صلب الجدول هو المساحة A تحت المنحني الطبيعي المعياري من ∞ - إلى $Z(A)$ مئينات مختار الاحتمال المتجمع A



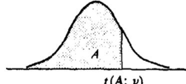
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

مئينات مختارة

الاحتمال المجتمع							
A:	.90	.95	.975	.98	.99	.995	.999
z(A):	1.282	1.645	1.960	2.054	2.326	2.576	3.090

جدول (أ ـ ٢) مئينات التوزيع t

العدد في صلب الجدول هو $t(A;\nu)$ حيث $P\{t(\nu) \leq t(A;\nu)\} = A$



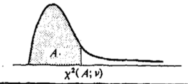
	A						
ν	.60	.70	.80	.85	.90	.95	.975
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

تتمة جدول (أ-٢) مئينات التوزيع t

ν	A .98	.985	.99	.9925	.995	.9975	.9995
1	15.895	21.205	31.821	42.434	63.657	127.322	636.590
2	4.849	5.643	6.965	8.073	9.925	14.089	31.598
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.849
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.690
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.659
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.291

جدول (ا-٣) مئينات التوزيع χ^2

العدد في صلب الجدول هو $\chi^2(A;\nu)$ حيث $P\{\chi^2(\nu) \le \chi^2(A;\nu)\} = A$



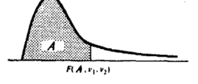
ν	A: .005	.010	.025	.050	.100	.900	.950	.975	.990	.995
1	$0.0^{4}393$	$0.0^{3}157$	$0.0^{3}982$	$0.0^{2}393$	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

المصدر: Reprinted, with permission, from C.M. Thompson. "Table of Percentage Points of the Chi-Square Distribution." *Biometrika* 32 (1941), pp. 188-89.

جدول (ا-٤) مئينات التوزيع F

العدد في صلب الجدول $F(A\,;\,\nu_1\,,\,\nu_2)$ حيث $P\{F(\nu_1\,,\,\nu_2) \le F\,(A;\,\nu_1\,,\,\nu_2)\} = A$

مئينات التوزيع F



$$F(A;\nu_1,\nu_2)=\frac{1}{F(1-A;\nu_2,\nu_1)}$$

تتمة جدول (ا-٤) مئينات التوزيع F

د.ح المقام	A	د.ح البسط 1	2	3	4	5	6	7	8	9
1	.50	1.00	1.50	1.71	1.82	1.89	1.94	1.98	2.00	2.03
	.90	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9
	.95	161	200	216	225	230	234	237	239	241
	.975	648	800	864	900	922	937	948	957	963
	.99	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,981	6,022
	.995	16,211	20,000	21,615	22,500	23,056	23,437	23,715	23,925	24,091
	.999	405,280	500,000	540,380	562,500	576,400	585,940	592,870	598,140	602,280
2	.50	0.667	1.00	1.13	1.21	1.25	1.28	1.30	1.32	1.33
	.90	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.95	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4
	.975	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4
	.99	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4
	.995	199	199	199	199	199	199	199	199	199
	.999	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4
3	.50	0.585	0.881	1.00	1.06	1.10	1.13	1.15	1.16	1.17
	.90	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.95	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.975	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5
	.99	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3
	.995	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9
	.999	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9
4	.50	0.549	0.828	0.941	1.00	1.04	1.06	1.08	1.09	1.10
	.90	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.975	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.99	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7
	.995	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1
	.999	74.1	61.2	56.2	53.4	51.7	50.5	49.7	49.0	48.5
5	.50	0.528	0.799	0.907	0.965	1.00	1.02	1.04	1.05	1.06
	.90	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.95	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.975	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.99	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2
	.995	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8
	.999	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2
6	.50	0.515	0.780	0.886	0.942	0.977	1.00	1.02	1.03	1.04
	.90	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.95	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.975	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.99	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.995	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4
	.999	35.5	27.0	23.7	21.9	20.8	20.0	19.5	19.0	18.7
7	.50	0.506	0.767	0.871	0.926	0.960	0.983	1.00	1.01	1.02
	.90	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.95	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.975	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.99	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.995	16.2	12.4	10.9	10.1	9.52	9.16	8.89	8.68	8.51
	.999	29.2	21.7	18.8	17.2	16.2	15.5	15.0	14.6	14.3

تتمة جدول (أ-٤) مئينات التوزيع F

د.ح المقام	A	د.ح البسط								
		10	12	15	20	24	30	60	120	∞
1	.50	2.04	2.07	2.09	2.12	2.13	2.15	2.17	2.18	2.20
	.90	60.2	60.7	61.2	61.7	62.0	62.3	62.8	63.1	63.3
	.95	242	244	246	248	249	250	252	253	254
	.975	969	977	985	993	997	1,001	1,010	1,014	1,018
	.99	6,056	6,106	6,157	6,209	6,235	6,261	6,313	6,339	6,366
	.995	24,224	24,426	24,630	24,836	24,940	25,044	25,253	25,359	25,464
	.999	605,620	610,670	615,760	620,910	623,500	626,100	631,340	633,970	636,620
2	.50	1.34	1.36	1.38	1.39	1.40	1.41	1.43	1.43	1.44
	.90	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.48	9.49
	.95	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
	.975	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5
	.99	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5
	.995	199	199	199	199	199	199	199	199	200
	.999	999.4	999.4	999.4	999.4	999.5	999.5	999.5	999.5	999.5
3	.50	1.18	1.20	1.21	1.23	1.23	1.24	1.25	1.26	1.27
	.90	5.23	5.22	5.20	5.18	5.18	5.17	5.15	5.14	5.13
	.95	8.79	8.74	8.70	8.66	8.64	8.62	8.57	8.55	8.53
	.975	14.4	14.3	14.3	14.2	14.1	14.1	14.0	13.9	13.9
	.99	27.2	27.1	26.9	26.7	26.6	26.5	26.3	26.2	26.1
	.995	43.7	43.4	43.1	42.8	42.6	42.5	42.1	42.0	41.8
	.999	129.2	128.3	127.4	126.4	125.9	125.4	124.5	124.0	123.5
4	.50	1.11	1.13	1.14	1.15	1.16	1.16	1.18	1.18	1.19
	.90	3.92	3.90	3.87	3.84	3.83	3.82	3.79	3.78	3.76
	.95	5.96	5.91	5.86	5.80	5.77	5.75	5.69	5.66	5.63
	.975	8.84	8.75	8.66	8.56	8.51	8.46	8.36	8.31	8.26
	.99	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.6	13.5
	.995	21.0	20.7	20.4	20.2	20.0	19.9	19.6	19.5	19.3
	.999	48.1	47.4	46.8	46.1	45.8	45.4	44.7	44.4	44.1
5	.50	1.07	1.09	1.10	1.11	1.12	1.12	1.14	1.14	1.15
	.90	3.30	3.27	3.24	3.21	3.19	3.17	3.14	3.12	3.11
	.95	4.74	4.68	4.62	4.56	4.53	4.50	4.43	4.40	4.37
	.975	6.62	6.52	6.43	6.33	6.28	6.23	6.12	6.07	6.02
	.99	10.1	9.89	9.72	9.55	9.47	9.38	9.20	9.11	9.02
	.995	13.6	13.4	13.1	12.9	12.8	12.7	12.4	12.3	12.1
	.999	26.9	26.4	25.9	25.4	25.1	24.9	24.3	24.1	23.8
6	.50	1.05	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1.12
	.90	2.94	2.90	2.87	2.84	2.82	2.80	2.76	2.74	2.72
	.95	4.06	4.00	3.94	3.87	3.84	3.81	3.74	3.70	3.67
	.975	5.46	5.37	5.27	5.17	5.12	5.07	4.96	4.90	4.85
	.99	7.87	7.72	7.56	7.40	7.31	7.23	7.06	6.97	6.88
	.995	10.2	10.0	9.81	9.59	9.47	9.36	9.12	9.00	8.88
	.999	18.4	18.0	17.6	17.1	16.9	16.7	16.2	16.0	15.7
7	.50	1.03	1.04	1.05	1.07	1.07	1.08	1.09	1.10	1.10
	.90	2.70	2.67	2.63	2.59	2.58	2.56	2.51	2.49	2.47
	.95	3.64	3.57	3.51	3.44	3.41	3.38	3.30	3.27	3.23
	.975	4.76	4.67	4.57	4.47	4.42	4.36	4.25	4.20	4.14
	.99	6.62	6.47	6.31	6.16	6.07	5.99	5.82	5.74	5.65
	.995	8.38	8.18	7.97	7.75	7.65	7.53	7.31	7.19	7.08
	.999	14.1	13.7	13.3	12.9	12.7	12.5	12.1	11.9	11.7

تتمة جدول (ا-٤) مئينات التوزيع F

د.ح المقام	A	د.ح البسط								
		1	2	3	4	5	6	7	8	9
8	.50	0.499	0.757	0.860	0.915	0.948	0.971	0.988	1.00	1.01
	.90	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.975	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.99	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.995	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34
	.999	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12.0	11.8
9	.50	0.494	0.749	0.852	0.906	0.939	0.962	0.978	0.990	1.00
	.90	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.995	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54
	.999	22.9	16.4	13.9	12.6	11.7	11.1	10.7	10.4	10.1
10	.50	0.490	0.743	0.845	0.899	0.932	0.954	0.971	0.983	0.992
	.90	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.975	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.995	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
	.999	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96
12	.50	0.484	0.735	0.835	0.888	0.921	0.943	0.959	0.972	0.981
	.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.975	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.99	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.995	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
	.999	18.6	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.48
15	.50	0.478	0.726	0.826	0.878	0.911	0.933	0.949	0.960	0.970
	.90	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	.95	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	.975	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
	.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	.995	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
	.999	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26
20	.50	0.472	0.718	0.816	0.868	0.900	0.922	0.938	0.950	0.959
	.90	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	.95	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	.975	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	.995	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
	.999	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24
24	.50	0.469	0.714	0.812	0.863	0.895	0.917	0.932	0.944	0.953
	.90	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	.95	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	.975	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	.99	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	.995	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
	.999	14.0	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80

تتمة جدول (ا–٤) مئينات التوزيع F

د.ح المقام	A	د.ح البسط 10	12	15	20	24	30	60	120	∞
8	.50	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.08	1.09
	.90	2.54	2.50	2.46	2.42	2.40	2.38	2.34	2.32	2.29
	.95	3.35	3.28	3.22	3.15	3.12	3.08	3.01	2.97	2.93
	.975	4.30	4.20	4.10	4.00	3.95	3.89	3.78	3.73	3.67
	.99	5.81	5.67	5.52	5.36	5.28	5.20	5.03	4.95	4.86
	.995	7.21	7.01	6.81	6.61	6.50	6.40	6.18	6.06	5.95
	.999	11.5	11.2	10.8	10.5	10.3	10.1	9.73	9.53	9.33
9	.50	1.01	1.02	1.03	1.04	1.05	1.05	1.07	1.07	1.08
	.90	2.42	2.38	2.34	2.30	2.28	2.25	2.21	2.18	2.16
	.95	3.14	3.07	3.01	2.94	2.90	2.86	2.79	2.75	2.71
	.975	3.96	3.87	3.77	3.67	3.61	3.56	3.45	3.39	3.33
	.99	5.26	5.11	4.96	4.81	4.73	4.65	4.48	4.40	4.31
	.995	6.42	6.23	6.03	5.83	5.73	5.62	5.41	5.30	5.19
	.999	9.89	9.57	9.24	8.90	8.72	8.55	8.19	8.00	7.81
10	.50	1.00	1.01	1.02	1.03	1.04	1.05	1.06	1.06	1.07
	.90	2.32	2.28	2.24	2.20	2.18	2.16	2.11	2.08	2.06
	.95	2.98	2.91	2.84	2.77	2.74	2.70	2.62	2.58	2.54
	.975	3.72	3.62	3.52	3.42	3.37	3.31	3.20	3.14	3.08
	.99	4.85	4.71	4.56	4.41	4.33	4.25	4.08	4.00	3.91
	.995	5.85	5.66	5.47	5.27	5.17	5.07	4.86	4.75	4.64
	.999	8.75	8.45	8.13	7.80	7.64	7.47	7.12	6.94	6.76
12	.50	0.989	1.00	1.01	1.02	1.03	1.03	1.05	1.05	1.06
	.90	2.19	2.15	2.10	2.06	2.04	2.01	1.96	1.93	1.90
	.95	2.75	2.69	2.62	2.54	2.51	2.47	2.38	2.34	2.30
	.975	3.37	3.28	3.18	3.07	3.02	2.96	2.85	2.79	2.72
	.99	4.30	4.16	4.01	3.86	3.78	3.70	3.54	3.45	3.36
	.995	5.09	4.91	4.72	4.53	4.43	4.33	4.12	4.01	3.90
	.999	7.29	7.00	6.71	6.40	6.25	6.09	5.76	5.59	5.42
15	.50	0.977	0.989	1.00	1.01	1.02	1.02	1.03	1.04	1.05
	.90	2.06	2.02	1.97	1.92	1.90	1.87	1.82	1.79	1.76
	.95	2.54	2.48	2.40	2.33	2.29	2.25	2.16	2.11	2.07
	.975	3.06	2.96	2.86	2.76	2.70	2.64	2.52	2.46	2.40
	.99	3.80	3.67	3.52	3.37	3.29	3.21	3.05	2.96	2.87
	.995	4.42	4.25	4.07	3.88	3.79	3.69	3.48	3.37	3.26
	.999	6.08	5.81	5.54	5.25	5.10	4.95	4.64	4.48	4.31
20	.50	0.966	0.977	0.989	1.00	1.01	1.01	1.02	1.03	1.03
	.90	1.94	1.89	1.84	1.79	1.77	1.74	1.68	1.64	1.61
	.95	2.35	2.28	2.20	2.12	2.08	2.04	1.95	1.90	1.84
	.975	2.77	2.68	2.57	2.46	2.41	2.35	2.22	2.16	2.09
	.99	3.37	3.23	3.09	2.94	2.86	2.78	2.61	2.52	2.42
	.995	3.85	3.68	3.50	3.32	3.22	3.12	2.92	2.81	2.69
	.999	5.08	4.82	4.56	4.29	4.15	4.00	3.70	3.54	3.38
24	.50	0.961	0.972	0.983	0.994	1.00	1.01	1.02	1.02	1.03
	.90	1.88	1.83	1.78	1.73	1.70	1.67	1.61	1.57	1.53
	.95	2.25	2.18	2.11	2.03	1.98	1.94	1.84	1.79	1.73
	.975	2.64	2.54	2.44	2.33	2.27	2.21	2.08	2.01	1.94
	.99	3.17	3.03	2.89	2.74	2.66	2.58	2.40	2.31	2.21
	.995	3.59	3.42	3.25	3.06	2.97	2.87	2.66	2.55	2.43
	.999	4.64	4.39	4.14	3.87	3.74	3.59	3.29	3.14	2.97

تتمة جدول (ا-٤) مئينات التوزيع F

د.ح المقام	A	د.ح البسط 1	2	3	4	5	6	7	8	9
30	.50	0.466	0.709	0.807	0.858	0.890	0.912	0.927	0.939	0.948
	.90	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	.975	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
	.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	.995	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
	.999	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
60	.50	0.461	0.701	0.798	0.849	0.880	0.901	0.917	0.928	0.937
	.90	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
	.95	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	.975	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
	.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
	.995	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01
	.999	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69
120	.50	0.458	0.697	0.793	0.844	0.875	0.896	0.912	0.923	0.932
	.90	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
	.95	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
	.975	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
	.99	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
	.995	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
	.999	11.4	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38
∞	.50	0.455	0.693	0.789	0.839	0.870	0.891	0.907	0.918	0.927
	.90	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63
	.95	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88
	.975	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11
	.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41
	.995	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62
	.999	10.8	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10

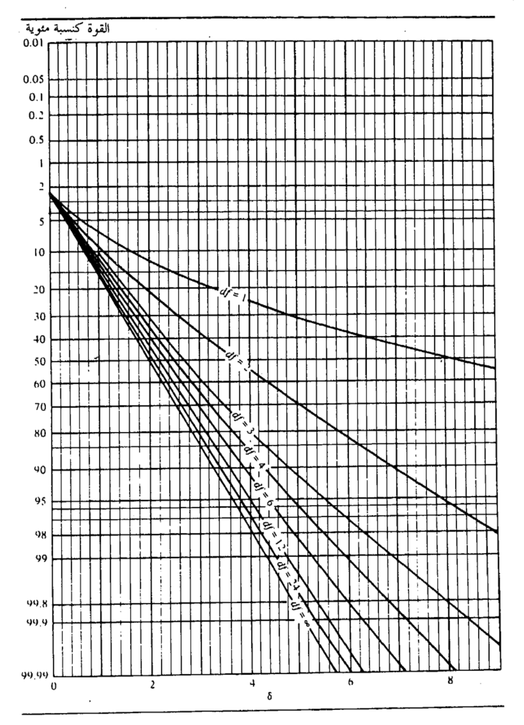
تتمة جدول (ا-٤) مئينات التوزيع F

د.ح المقام	A	10	12	15	20	24	30	60	120	∞
		د.ح البسط								
30	.50	0.955	0.966	0.978	0.989	0.994	1.00	1.01	1.02	1.02
	.90	1.82	1.77	1.72	1.67	1.64	1.61	1.54	1.50	1.46
	.95	2.16	2.09	2.01	1.93	1.89	1.84	1.74	1.68	1.62
	.975	2.51	2.41	2.31	2.20	2.14	2.07	1.94	1.87	1.79
	.99	2.98	2.84	2.70	2.55	2.47	2.39	2.21	2.11	2.01
	.995	3.34	3.18	3.01	2.82	2.73	2.63	2.42	2.30	2.18
	.999	4.24	4.00	3.75	3.49	3.36	3.22	2.92	2.76	2.59
60	.50	0.945	0.956	0.967	0.978	0.983	0.989	1.00	1.01	1.01
	.90	1.71	1.66	1.60	1.54	1.51	1.48	1.40	1.35	1.29
	.95	1.99	1.92	1.84	1.75	1.70	1.65	1.53	1.47	1.39
	.975	2.27	2.17	2.06	1.94	1.88	1.82	1.67	1.58	1.48
	.99	2.63	2.50	2.35	2.20	2.12	2.03	1.84	1.73	1.60
	.995	2.90	2.74	2.57	2.39	2.29	2.19	1.96	1.83	1.69
	.999	3.54	3.32	3.08	2.83	2.69	2.55	2.25	2.08	1.89
120	.50	0.939	0.950	0.961	0.972	0.978	0.983	0.994	1.00	1.01
	.90	1.65	1.60	1.55	1.48	1.45	1.41	1.32	1.26	1.19
	.95	1.91	1.83	1.75	1.66	1.61	1.55	1.43	1.35	1.25
	.975	2.16	2.05	1.95	1.82	1.76	1.69	1.53	1.43	1.31
	.99	2.47	2.34	2.19	2.03	1.95	1.86	1.66	1.53	1.38
	.995	2.71	2.54	2.37	2.19	2.09	1.98	1.75	1.61	1.43
	.999	3.24	3.02	2.78	2.53	2.40	2.26	1.95	1.77	1.54
∞	.50	0.934	0.945	0.956	0.967	0.972	0.978	0.989	0.994	1.00
	.90	1.60	1.55	1.49	1.42	1.38	1.34	1.24	1.17	1.00
	.95	1.83	1.75	1.67	1.57	1.52	1.46	1.32	1.22	1.00
	.975	2.05	1.94	1.83	1.71	1.64	1.57	1.39	1.27	1.00
	.99	2.32	2.18	2.04	1.88	1.79	1.70	1.47	1.32	1.00
	.995	2.52	2.36	2.19	2.00	1.90	1.79	1.53	1.36	1.00
	.999	2.96	2.74	2.51	2.27	2.13	1.99	1.66	1.45	1.00

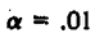
المصدر: Reprinted from Table 5 of Pearson and Hartley, *Biometrika Tables for Statisticians*, Volume 2, 1972, published by the Cambridge University Press, on behalf of The Biometria Society, by permission of the authors and publishers.

جدول (ا–٥) دالة القوة للاختبار t ذي الجانبين

$\alpha = .05$



تتمة جدول (ا-٥) دالة القوة للاختبار t ذي الجانبين



المصدر: Reprinted, with permission, from D.B. Owen, *Handbook of Statistical Tables* (Reading, Mass. Addison Wesley Publishing. 1962), pp. 32, 34. Courtesy of U.S. Atomic Energy Commission.

جدول (ا-٦) حدًا اختبار دوربين ـ واتسون

مستوى الأهمية $\alpha = 0.05$

n	$p-1=1$		$p-1=2$		$p-1=3$		$p-1=4$		$p-1=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

تتمة جدول (ا-٦) حدّا اختبار دوربين - واتسون

مستوى الأهمية $\alpha = 0.01$

n	$p-1=1$		$p-1=2$		$p-1=3$		$p-1=4$		$p-1=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

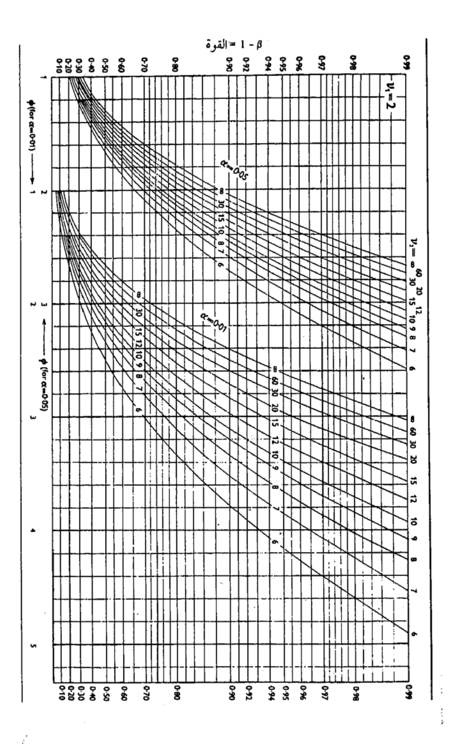
المصدر: Reprinted, with permission, from J. Durbin and G.S. Watson, "Testing fro Serial Correlation in Least Squares Regression. II", *Biometrika* 38 (1951), pp. 159-78.

جدول (ا-٧) جدول التحويل z' لمعامل الارتباط

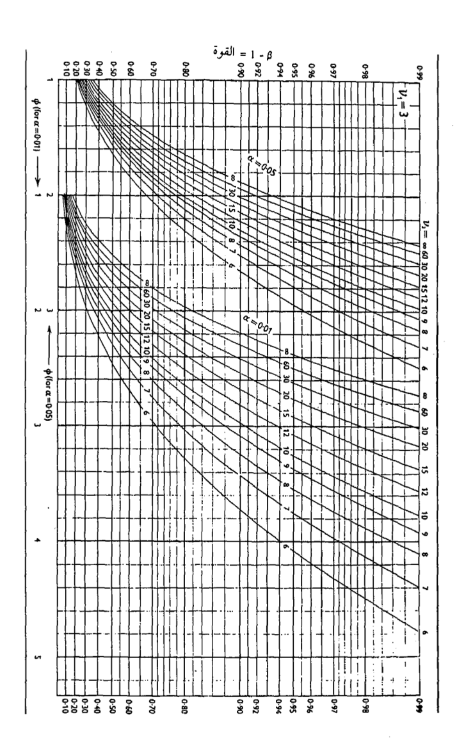
r ρ	z' ζ	r ρ	z' ζ	r ρ	z' ζ	r ρ	z' ζ
.00	.0000	.25	.2554	.50	.5493	.75	.973
.01	.0100	.26	.2661	.51	.5627	.76	.996
.02	.0200	.27	.2769	.52	.5763	.77	1.020
.03	.0300	.28	.2877	.53	.5901	.78	1.045
.04	.0400	.29	.2986	.54	.6042	.79	1.071
.05	.0500	.30	.3095	.55	.6184	.80	1.099
.06	.0601	.31	.3205	.56	.6328	.81	1.127
.07	.0701	.32	.3316	.57	.6475	.82	1.157
.08	.0802	.33	.3428	.58	.6625	.83	1.188
.09	.0902	.34	.3541	.59	.6777	.84	1.221
.10	.1003	.35	.3654	.60	.6931	.85	1.256
.11	.1104	.36	.3769	.61	.7089	.86	1.293
.12	.1206	.37	.3884	.62	.7250	.87	1.333
.13	.1307	.38	.4001	.63	.7414	.88	1.376
.14	.1409	.39	.4118	.64	.7582	.89	1.422
.15	.1511	.40	.4236	.65	.7753	.90	1.472
.16	.1614	.41	.4356	.66	.7928	.91	1.528
.17	.1717	.42	.4477	.67	.8107	.92	1.589
.18	.1820	.43	.4599	.68	.8291	.93	1.658
.19	.1923	.44	.4722	.69	.8480	.94	1.738
.20	.2027	.45	.4847	.70	.8673	.95	1.832
.21	.2132	.46	.4973	.71	.8872	.96	1.946
.22	.2237	.47	.5101	.72	.9076	.97	2.092
.23	.2342	.48	.5230	.73	.9287	.98	2.298
.24	.2448	.49	.5361	.74	.9505	.99	2.647

المصدر: Abridged from Table 14 of Pearson and Hartley, *Biometrika Tables for Statisticians*, volume 1, 1966, published by the Cambridge University Press, on behalf of The Biometrika Society, by permission of the authors and publishers.

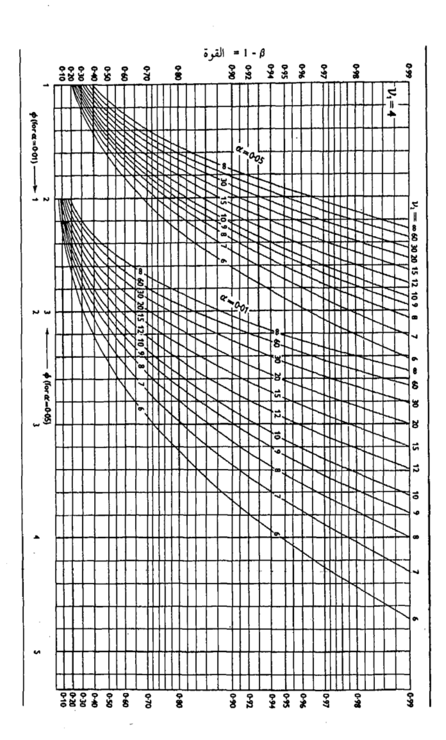
جدول ا - ٨) دالة القوة لتحليل تباين (نموذج تأثيرات مثبتة لمستويات عامل)



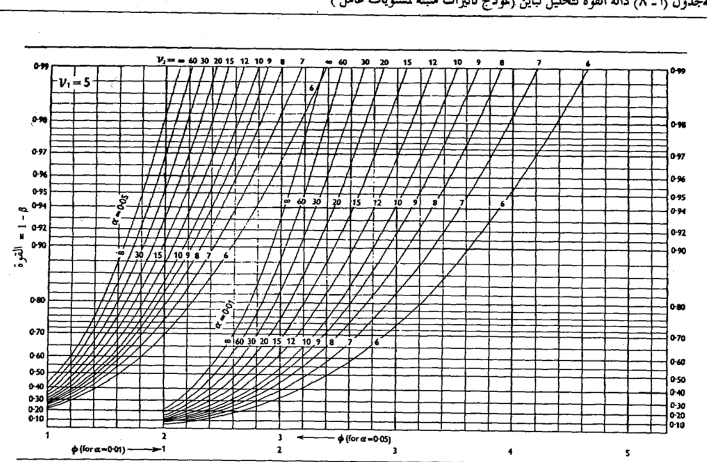
تتمة جدول (١-٨) دالة القوة لتحليل تباين (نموذج تأثيرات مثبتة لمستويات عامل)



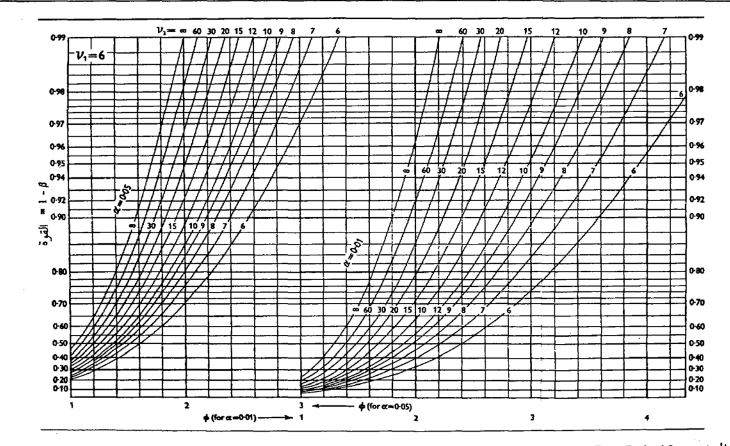
تتمة جدول (أ - ٨) دالة القوة لتحليل تباين (نموذج تأثيرات مثبتة لمستويات عامل)



تتمة جدول (أ - ٨) دالة القوة لتحليل تباين (نموذج تأثيرات مثبتة لمستويات عامل)



تتمة جدول (ا ـ ٨) دالة القوة لتحليل تباين (نموذج تأثيرات مثبتة لمستويات عامل)



المصدر: Reprinted, with permission, from E.S. Pearson and H.O. Harly, "Charts of the Power Function for Analysis of Variance Tests, Derived from the Non-Central *F*-Distribution," *Biometrika* 38 (1951), pp. 112-30.

جدول (ا ـ ٩) مئينات توزيع المدى المعيّر تقديرا

العدد في صُلب الجدول هو $q(1-\alpha; r, \nu)$ حيث $P\{q(r,\nu) \le q(1-\alpha; r, \nu)\} = 1-\alpha$

$1-\alpha = .90$

ν	r = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	8.93	13.4	16.4	18.5	20.2	21.5	22.6	23.6	24.5	25.2	25.9	26.5	27.1	27.6	28.1	28.5	29.0	29.3	29.7
2	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9.41	9.72	10.0	10.3	10.5	10.7	10.9	11.1	11.2	11.4	11.5	11.7
3	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29	7.49	7.67	7.83	7.98	8.12	8.25	8.37	8.48	8.58	8.68
4	3.01	3.98	4.59	5.03	5.39	5.68	5.93	6.14	6.33	6.49	6.65	6.78	6.91	7.02	7.13	7.23	7.33	7.41	7.50
5	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82	5.97	6.10	6.22	6.34	6.44	6.54	6.63	6.71	6.79	6.86
6	2.75	3.56	4.07	4.44	4.73	4.97	5.17	5.34	5.50	5.64	5.76	5.87	5.98	6.07	6.16	6.25	6.32	6.40	6.47
7	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28	5.41	5.53	5.64	5.74	5.83	5.91	5.99	6.06	6.13	6.19
8	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.80	5.87	5.93	6.00
9	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01	5.13	5.23	5.33	5.42	5.51	5.58	5.66	5.72	5.79	5.85
10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91	5.03	5.13	5.23	5.32	5.40	5.47	5.54	5.61	5.67	5.73
11	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84	4.95	5.05	5.15	5.23	5.31	5.38	5.45	5.51	5.57	5.63
12	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78	4.89	4.99	5.08	5.16	5.24	5.31	5.37	5.44	5.49	5.55
13	2.50	3.18	3.59	3.88	4.12	4.30	4.46	4.60	4.72	4.83	4.93	5.02	5.10	5.18	5.25	5.31	5.37	5.43	5.48
14	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68	4.79	4.88	4.97	5.05	5.12	5.19	5.26	5.32	5.37	5.43
15	2.48	3.14	3.54	3.83	4.05	4.23	4.39	4.52	4.64	4.75	4.84	4.93	5.01	5.08	5.15	5.21	5.27	5.32	5.38
16	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61	4.71	4.81	4.89	4.97	5.04	5.11	5.17	5.23	5.28	5.33
17	2.46	3.11	3.50	3.78	4.00	4.18	4.33	4.46	4.58	4.68	4.77	4.86	4.93	5.01	5.07	5.13	5.19	5.24	5.30
18	2.45	3.10	3.49	3.77	[illegible]	4.16	4.31	4.44	4.55	4.65	4.75	4.83	4.90	4.98	5.04	5.10	5.16	5.21	5.26
19	2.45	3.09	3.47	3.75	3.97	4.14	4.29	4.42	4.53	4.63	4.72	4.80	4.88	4.95	5.01	5.07	5.13	5.18	5.23
20	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51	4.61	4.70	4.78	4.85	4.92	4.99	5.05	5.10	5.16	5.20
24	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44	4.54	4.63	4.71	4.78	4.85	4.91	4.97	5.02	5.07	5.12
30	2.40	3.02	3.39	3.65	3.85	4.02	4.16	4.28	4.38	4.47	4.56	4.64	4.71	4.77	4.83	4.89	4.94	4.99	5.03
40	2.38	2.99	3.35	3.60	3.80	3.96	4.10	4.21	4.32	4.41	4.49	4.56	4.63	4.69	4.75	4.81	4.86	4.90	4.95
60	2.36	2.96	3.31	3.56	3.75	3.91	4.04	4.16	4.25	4.34	4.42	4.49	4.56	4.62	4.67	4.73	4.78	4.82	4.86
120	2.34	2.93	3.28	3.52	3.71	3.86	3.99	4.10	4.19	4.28	4.35	4.42	4.48	4.54	4.60	4.65	4.69	4.74	4.78
∞	2.33	2.90	3.24	3.48	3.66	3.81	3.93	4.04	4.13	4.21	4.28	4.35	4.41	4.47	4.52	4.57	4.61	4.65	4.69

تتمة جدول (ا ـ ٩) مئينات توزيع المدى المعيّر تقديرا

$1 - \alpha = .95$

ν	r: 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18.0	27.0	32.8	37.1	40.4	43.1	45.4	47.4	49.1	50.6	52.0	53.2	54.3	55.4	56.3	57.2	58.0	58.8	59.6
2	6.08	8.33	9.80	10.9	11.7	12.4	13.0	13.5	14.0	14.4	14.7	15.1	15.4	15.7	15.9	16.1	16.4	16.6	16.8
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.2	10.3	10.5	10.7	10.8	11.0	11.1	11.2
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

تتمة جدول (ا - ٩) مئينات توزيع المدى المعيّر تقديرا

$1 - \alpha = .99$

ν	r = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90.0	135	164	186	202	216	227	237	246	253	260	266	272	277	282	286	290	294	298
2	14.0	19.0	22.3	24.7	26.6	28.2	29.5	30.7	31.7	32.6	33.4	34.1	34.8	35.4	36.0	36.5	37.0	37.5	37.9
3	8.26	10.6	12.2	13.3	14.2	15.0	15.6	16.2	16.7	17.1	17.5	17.9	18.2	18.5	18.8	19.1	19.3	19.5	19.8
4	6.51	8.12	9.17	9.96	10.6	11.1	11.5	11.9	12.3	12.6	12.8	13.1	13.3	13.5	13.7	13.9	14.1	14.2	14.4
5	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.2	10.5	10.7	10.9	11.1	11.2	11.4	11.6	11.7	11.8	11.9
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.49	9.65	9.81	9.95	10.1	10.2	10.3	10.4	10.5
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49	7.65	7.78	7.91	8.03	8.13	8.23	8.32	8.41	8.49	8.57
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.48	7.60	7.71	7.81	7.91	7.99	8.07	8.15	8.22
11	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90	7.01	7.10	7.19	7.27	7.34	7.42	7.48	7.55
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.12	7.20	7.27	7.33	7.39
15	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.80	6.87	6.94	7.00	7.05
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.72	6.79	6.85	6.91	6.96
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.29	6.37	6.45	6.52	6.59	6.65	6.71	6.76	6.82
24	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60	5.69	5.77	5.84	5.90	5.96	6.02	6.07	6.12	6.17	6.21
60	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.79	5.84	5.89	5.93	5.98	6.02
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.38	5.44	5.51	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

المصدر: Reprinted, with permission, from Henry Scheffé. *The Analysis of Variance* (New York: John Wiley & Sons, 1959), pp. 434-36.

جدول (ا - ١٠) جدول تحديد حجم العينة في تجربة تحليل تباين (نموذج تأثيرات مثبتة لمستويات عامل)

القوة $1 - \beta = .70$

	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$				$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
	α				α				α				α				α				α				α			
r	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	7	11	14	21	5	7	9	15	4	6	7	11	3	4	6	9	3	4	5	7	2	3	4	5	2	3	3	5
3	9	13	17	25	6	9	11	17	5	7	8	12	4	5	7	10	3	4	5	8	3	3	4	6	2	3	3	5
4	11	15	19	28	7	10	13	19	5	7	9	13	4	6	7	10	4	5	6	8	3	4	4	6	2	3	4	5
5	12	17	21	30	8	11	14	20	6	8	10	14	5	6	8	11	4	5	6	9	3	4	5	6	3	3	4	5
6	13	18	22	32	9	12	15	21	6	9	11	15	5	7	8	12	4	5	7	9	3	4	5	7	3	3	4	5
7	14	19	24	34	9	13	16	22	7	9	11	16	5	7	9	12	4	6	7	10	3	4	5	7	3	3	4	5
8	15	20	25	35	10	13	16	23	7	10	12	17	6	7	9	13	5	6	7	10	3	4	5	7	3	3	4	5
9	15	21	26	37	10	14	17	24	7	10	12	17	6	8	9	13	5	6	8	10	3	4	5	7	3	4	4	6
10	16	22	27	38	11	14	18	25	8	10	13	18	6	8	10	14	5	6	8	11	4	5	6	7	3	4	4	6

القوة $1 - \beta = .80$

	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$				$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
	α				α				α				α				α				α				α			
r	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	10	14	17	26	7	9	12	17	5	7	9	13	4	5	7	10	3	4	6	8	3	3	4	6	2	3	4	5
3	12	17	21	30	8	11	14	20	6	8	10	14	5	6	8	11	4	5	6	9	3	4	5	7	3	3	4	5
4	14	19	23	33	9	13	15	22	7	9	11	16	5	7	9	12	4	6	7	10	3	4	5	7	3	3	4	5
5	16	21	25	35	10	14	17	23	8	10	12	17	6	8	9	13	5	6	7	10	4	4	5	7	3	4	4	6
6	17	22	27	38	11	15	18	25	8	11	13	18	6	8	10	13	5	7	8	11	4	5	6	8	3	4	4	6
7	18	24	29	39	12	16	19	26	9	11	14	18	7	9	10	14	5	7	8	11	4	5	6	8	3	4	5	6
8	19	25	30	41	12	16	20	27	9	12	14	19	7	9	11	15	6	7	9	12	4	5	6	8	3	4	5	6
9	20	26	31	43	13	17	21	28	9	12	15	20	7	9	11	15	6	7	9	12	4	5	6	8	3	4	5	6
10	21	27	33	44	14	18	21	29	10	13	15	21	8	10	12	16	6	8	9	12	4	5	6	8	3	4	5	6

تتمة جدول (ا - ١٠) جدول تحديد حجم العينة في تجربة تحليل تباين (نموذج تأثيرات مثبتة لمستويات عامل)

القوة $1 - \beta = .90$

	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$				$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
	α				α				α				α				α				α				α			
r	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	14	18	23	32	9	12	15	21	7	9	11	15	5	7	8	12	4	6	7	10	3	4	5	7	3	3	4	6
3	17	22	27	37	11	15	18	24	8	11	13	18	6	8	10	13	5	7	8	11	4	5	6	8	3	4	5	6
4	20	25	30	40	13	16	20	27	9	12	14	19	7	9	11	15	6	7	9	12	4	5	6	8	3	4	5	6
5	21	27	32	43	14	18	21	28	10	13	15	20	8	10	12	15	6	8	9	12	4	5	6	9	4	4	5	7
6	22	29	34	46	15	19	23	30	11	14	16	21	8	10	12	16	7	8	10	13	5	6	7	9	4	4	5	7
7	24	31	36	48	16	20	24	31	11	14	17	22	9	11	13	17	7	9	10	13	5	6	7	9	4	5	5	7
8	26	32	38	50	17	21	25	33	12	15	18	23	9	11	13	17	7	9	11	14	5	6	7	9	4	5	6	7
9	27	33	40	52	17	22	26	34	13	16	18	24	9	12	14	18	8	9	11	14	5	6	8	10	4	5	6	7
10	28	35	41	54	18	23	27	35	13	16	19	25	10	12	14	19	8	10	11	15	5	7	8	10	4	5	6	7

القوة $1 - \beta = .95$

	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$				$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
	α				α				α				α				α				α				α			
r	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	18	23	27	38	12	15	18	25	9	11	13	18	7	8	10	14	5	7	8	11	4	5	6	8	3	4	5	6
3	22	27	32	43	14	18	21	29	10	13	15	20	8	10	12	16	6	8	9	12	5	6	7	9	4	4	5	7
4	25	30	36	47	16	20	23	31	12	14	17	22	9	11	13	17	7	9	10	13	5	6	7	9	4	5	5	7
5	27	33	39	51	18	22	25	33	13	15	18	23	10	12	14	18	8	9	11	14	5	6	7	10	4	5	6	7
6	29	35	41	53	19	23	27	35	13	16	19	25	10	12	14	19	8	10	11	15	6	7	8	10	4	5	6	8
7	30	37	43	56	20	24	28	36	14	17	20	26	11	13	15	19	8	10	12	15	6	7	8	10	4	5	6	8
8	32	39	45	58	21	25	29	38	15	18	21	27	11	14	16	20	9	11	12	16	6	7	8	11	5	5	6	8
9	33	40	47	60	22	26	30	39	15	19	22	28	12	14	16	21	9	11	13	16	6	8	9	11	5	6	6	8
10	34	42	48	62	22	27	31	40	16	19	22	29	12	15	17	21	9	11	13	17	6	8	9	11	5	6	7	8

المصدر: Reprinted, with permission, from T.L. Bratcher, M.A. Moran, and W.J. Zimmer, "Tables of Sample Sizes in the Analysis of Variance," *Journal of Quality Technology* 2 (1970), pp. 156-64. Copyright American Society for Quality Control, Inc

جدول (ا - ١١) جدول $\lambda\sqrt{n}/\sigma$ لتحديد حجم العينة من أجل إيجاد "أفضل" متوسط بين متوسطات r من المجتمعات.

عدد المجتمعات	احتمال التحديد الصحيح $(1 - \alpha)$، عدد المجتمعات r،		
	.90	.95	.99
2	1.8124	2.3262	3.2900
3	2.2302	2.7101	3.6173
4	2.4516	2.9162	3.7970
5	2.5997	3.0552	3.9196
6	2.7100	3.1591	4.0121
7	2.7972	3.2417	4.0861
8	2.8691	3.3099	4.1475
9	2.9301	3.3679	4.1999
10	2.9829	3.4182	4.2456

المصدر: Reprinted, with permission, from R.E. Bechhofer, "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances, "*The Annals of Machematical Statistics* 25 (1954), pp. 16-39.

جدول (ا - ١٢) مئينات توزيع الاحصاءة H

العدد في صلب الجدول هو $H(1 - \alpha; r, df)$ حيث : $P\{H \leq H(1 - \alpha; r, df)\} = 1 - \alpha$

$1 - \alpha = .95$

	r										
df	2	3	4	5	6	7	8	9	10	11	12
2	39.0	87.5	142	202	266	333	403	475	550	626	704
3	15.4	27.8	39.2	50.7	62.0	72.9	83.5	93.9	104	114	124
4	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.1	44.6	48.0	51.4
5	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
6	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
7	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
8	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
9	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
10	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
12	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
15	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
20	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
30	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
60	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

$1 - \alpha = .99$

	r										
df	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1,036	1,362	1,705	2,063	2,432	2,813	3,204	3,605
3	47.5	85	120	151	184	216	249	281	310	337	361
4	23.2	37	49	59	69	79	89	97	106	113	120
5	14.9	22	28	33	38	42	46	50	54	57	60
6	11.1	15.5	19.1	22	25	27	30	32	34	36	37
7	8.89	12.1	14.5	16.5	18.4	20	22	23	24	26	27
8	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21
9	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
10	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
12	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
15	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
20	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.8	5.9
30	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2
60	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7
∞	1.00	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

المصدر: Reprinted, with permission, from H.A. David, "Upper 5 and 1% Points of the Maximum *F*-Ratio", *Biometrika* 39 (1952), pp. 422-24.

جدول (ا ـ ١٣) مربعات لاتينية قياسية مختارة

3 × 3

A	B	C
B	C	A
C	A	B

4 × 4

1

A	B	C	D
B	A	D	C
C	D	B	A
D	C	A	B

2

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

3

A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

4

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

5 × 5

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A

6 × 6

A	B	C	D	E	F
B	F	D	C	A	E
C	D	E	F	B	A
D	A	F	E	C	B
E	C	A	B	F	D
F	E	B	A	D	C

7 × 7

A	B	C	D	E	F	G
B	C	D	E	F	G	A
C	D	E	F	G	A	B
D	E	F	G	A	B	C
E	F	G	A	B	C	D
F	G	A	B	C	D	E
G	A	B	C	D	E	F

8 × 8

A	B	C	D	E	F	G	H
B	C	D	E	F	G	H	A
C	D	E	F	G	H	A	B
D	E	F	G	H	A	B	C
E	F	G	H	A	B	C	D
F	G	H	A	B	C	D	E
G	H	A	B	C	D	E	F
H	A	B	C	D	E	F	G

9 × 9

A	B	C	D	E	F	G	H	I
B	C	D	E	F	G	H	I	A
C	D	E	F	G	H	I	A	B
D	E	F	G	H	I	A	B	C
E	F	G	H	I	A	B	C	D
F	G	H	I	A	B	C	D	E
G	H	I	A	B	C	D	E	F
H	I	A	B	C	D	E	F	G
I	A	B	C	D	E	F	G	H

الملحق ب

مجموعات من البيانات

مجموعة بيانات (ب - ١) *SENIC*

الهدف الرئيس من دراسة فعالية التحكم بانتان مستشفى (مشروع **SENIC**) ه
تحديد ما إذا كانت برامج مسح وضبط الانتانات قد خفضت من معـدلات الانتانـار
المكتسبة من مستشفى في مستشفيات الولايات المتحدة. وتتألف مجموعة البيانات هـذ
من عينة عشوائية من 113 مستشفى اختيرت مـن بـين الـ 338 مستشـفى الـتي تناوله
المسح.

ويتضمن كل سطر من مجموعة البيانات رقم تسلسـل وأحـد عشـر متغـيرا لمستشـف
بمفرده. والبيانات المقدمة هنا خاصة بفترة الدراسة ١٩٧٦/٧٥م، والمتغيرات الـ ١٢ هي:

رقم المتغير	اسم المتغير	وصف المتغير
١	رقم تسلسل	1 - 113
٢	طول فترة الإقامة	متوسط فترة الإقامة لجميع المرضى في مستشفى (بالأيام)
٣	العمر	متوسط عمر المرضى (بالسنوات)
٤	مخاطرة الإصابة	تقدير لاحتمال اكتساب انتان في المستشـفى ؤ المتوسط (كنسبة مئوية)
٥	نسبة الزرع الروتيني	نسـبة عـدد حـالات الـزرع إلى عـدد المرضــى بدون إشارات أو أعـراض انتـان مكتسـب مـ المستشفى، مضروبة بمائة.
٦	نسبة تصوير الصدر الروتيني بأشعة X	نسبة عدد الصور التي تمـت بأشـعة X إلى عـد المرضى بـدون إشـارات أو أعـراض ذات الرئـ مضروبة بمائة.

رقم المتغير	اسم المتغير	وصف المتغير
٧	عدد الأسرّة	متوسط عدد الأسرة في المستشفى خلال فترة الدراسة
٨	الانتماء إلى مدرسة طبية	نعم = 1 لا = 2
٩	المنطقة	منطقة جغرافية حيث: $NE = 1$، $NC = 2$، $S = 3$ و $W = 4$
١٠	متوسط التعداد اليومي	متوسط عدد المرضى اليومي في مستشفى خلال فترة الدراسة
١١	عدد الممرضات	العدد المتوسط للممرضات التطبيقيات المجازات والمسجلات المتفرغات، خلال فترة الدراسة (عدد المتفرغات + نصف عدد المتفرغات جزئيا)
١٢	التسهيلات والخدمات اليومية	النسبة المئوية لـ 35 من التسهيلات والخدمات الممكنة التي يوفرها المستشفى

المصدر: Special Issue, "The SENIC Project". *American Journal of Epidemoiology* 111 (1980), pp. 465-653.

Data obtained from: Robert W. Haley, M.D., Hospital Infections Program, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333.

1	2	3	4	5	6	7	8	9	10	11	12
1	7.13	55.7	4.1	9.0	39.6	279	2	4	207	241	60.0
2	8.82	58.2	1.6	3.8	51.7	80	2	2	51	52	40.0
3	8.34	56.9	2.7	8.1	74.0	107	2	3	82	54	20.0
4	8.95	53.7	5.6	18.9	122.8	147	2	4	53	148	40.0
5	11.20	56.5	5.7	34.5	88.9	180	2	1	134	151	40.0
6	9.76	50.9	5.1	21.9	97.0	150	2	2	147	106	40.0
7	9.68	57.8	4.6	16.7	79.0	186	2	3	151	129	40.0
8	11.18	45.7	5.4	60.5	85.8	640	1	2	399	360	60.0
9	8.67	48.2	4.3	24.4	90.8	182	2	3	130	118	40.0
10	8.84	56.3	6.3	29.6	82.6	85	2	1	59	66	40.0
11	11.07	53.2	4.9	28.5	122.0	768	1	1	591	656	80.0
12	8.30	57.2	4.3	6.8	83.8	167	2	3	105	59	40.0
13	12.78	56.8	7.7	46.0	116.9	322	1	1	252	349	57.1
14	7.58	56.7	3.7	20.8	88.0	97	2	2	59	79	37.1
15	9.00	56.3	4.2	14.6	76.4	72	2	3	61	38	17.1
16	11.08	50.2	5.5	18.6	63.6	387	2	3	326	405	57.1
17	8.28	48.1	4.5	26.0	101.8	108	2	4	84	73	37.1
18	11.62	53.9	6.4	25.5	99.2	133	2	1	113	101	37.1
19	9.06	52.8	4.2	6.9	75.9	134	2	2	103	125	37.1
20	9.35	53.8	4.1	15.9	80.9	833	2	3	547	519	77.1
21	7.53	42.0	4.2	23.1	98.9	95	2	4	47	49	17.1
22	10.24	49.0	4.8	36.3	112.6	195	2	2	163	170	37.1
23	9.78	52.3	5.0	17.6	95.9	270	1	1	240	198	57.1
24	9.84	62.2	4.8	12.0	82.3	600	2	3	468	497	57.1
25	9.20	52.2	4.0	17.5	71.1	298	1	4	244	236	57.1
26	8.28	49.5	3.9	12.0	113.1	546	1	2	413	436	57.1
27	9.31	47.2	4.5	30.2	101.3	170	2	1	124	173	37.1
28	8.19	52.1	3.2	10.8	59.2	176	2	1	156	88	37.1
29	11.65	54.5	4.4	18.6	96.1	248	2	1	217	189	37.1
30	9.89	50.5	4.9	17.7	103.6	167	2	2	113	106	37.1
31	11.03	49.9	5.0	19.7	102.1	318	2	1	270	335	57.1
32	9.84	53.0	5.2	17.7	72.6	210	2	2	200	239	54.3
33	11.77	54.1	5.3	17.3	56.0	196	2	1	164	165	34.3
34	13.59	54.0	6.1	24.2	111.7	312	2	1	258	169	54.3
35	9.74	54.4	6.3	11.4	76.1	221	2	2	170	172	54.3
36	10.33	55.8	5.0	21.2	104.3	266	2	1	181	149	54.3
37	9.97	58.2	2.8	16.5	76.5	90	2	2	69	42	34.3
38	7.84	49.1	4.6	7.1	87.9	60	2	3	50	45	34.3
39	10.47	53.2	4.1	5.7	69.1	196	2	2	168	153	54.3
40	8.16	60.9	1.3	1.9	58.0	73	2	3	49	21	14.3
41	8.48	51.1	3.7	12.1	92.8	166	2	3	145	118	34.3
42	10.72	53.8	4.7	23.2	94.1	113	2	3	90	107	34.3
43	11.20	45.0	3.0	7.0	78.9	130	2	3	95	56	34.3
44	10.12	51.7	5.6	14.9	79.1	362	1	3	313	264	54.3
45	8.37	50.7	5.5	15.1	84.8	115	2	2	96	88	34.3
46	10.16	54.2	4.6	8.4	51.5	831	1	4	581	629	74.3
47	19.56	59.9	6.5	17.2	113.7	306	2	1	273	172	51.4
48	10.90	57.2	5.5	10.6	71.9	593	2	2	446	211	51.4
49	7.67	51.7	1.8	2.5	40.4	106	2	3	93	35	11.4
50	8.88	51.5	4.2	10.1	86.9	305	2	3	238	197	51.4
51	11.48	57.6	5.6	20.3	82.0	252	2	1	207	251	51.4
52	9.23	51.6	4.3	11.6	42.6	620	2	2	413	420	71.4
53	11.41	61.1	7.6	16.6	97.9	535	2	3	330	273	51.4
54	12.07	43.7	7.8	52.4	105.3	157	2	2	115	76	31.4
55	8.63	54.0	3.1	8.4	56.2	76	2	1	39	44	31.4
56	11.15	56.5	3.9	7.7	73.9	281	2	1	217	199	51.4

1	2	3	4	5	6	7	8	9	10	11	12
57	7.14	59.0	3.7	2.6	75.8	70	2	4	37	35	31.4
58	7.65	47.1	4.3	16.4	65.7	318	2	4	265	314	51.4
59	10.73	50.6	3.9	19.3	101.0	445	1	2	374	345	51.4
60	11.46	56.9	4.5	15.6	97.7	191	2	3	153	132	31.4
61	10.42	58.0	3.4	8.0	59.0	119	2	1	67	64	31.4
62	11.18	51.0	5.7	18.8	55.9	595	1	2	546	392	68.6
63	7.93	64.1	5.4	7.5	98.1	68	2	4	42	49	28.6
64	9.66	52.1	4.4	9.9	98.3	83	2	2	66	95	28.6
65	7.78	45.5	5.0	20.9	71.6	489	2	3	391	329	48.6
66	9.42	50.6	4.3	24.8	62.8	508	2	1	421	528	48.6
67	10.02	49.5	4.4	8.3	93.0	265	2	2	191	202	48.6
68	8.58	55.0	3.7	7.4	95.9	304	2	3	248	218	48.6
69	9.61	52.4	4.5	6.9	87.2	487	2	3	404	220	48.6
70	8.03	54.2	3.5	24.3	87.3	97	2	1	65	55	28.6
71	7.39	51.0	4.2	14.6	88.4	72	2	2	38	67	28.6
72	7.08	52.0	2.0	12.3	56.4	87	2	3	52	57	28.6
73	9.53	51.5	5.2	15.0	65.7	298	2	3	241	193	48.6
74	10.05	52.0	4.5	36.7	87.5	184	1	1	144	151	68.6
75	8.45	38.8	3.4	12.9	85.0	235	2	2	143	124	48.6
76	6.70	48.6	4.5	13.0	80.8	76	2	4	51	79	28.6
77	8.90	49.7	2.9	12.7	86.9	52	2	1	37	35	28.6
78	10.23	53.2	4.9	9.9	77.9	752	1	2	595	446	68.6
79	8.88	55.8	4.4	14.1	76.8	237	2	2	165	182	48.6
80	10.30	59.6	5.1	27.8	88.9	175	2	2	113	73	45.7
81	10.79	44.2	2.9	2.6	56.6	461	1	2	320	196	65.7
82	7.94	49.5	3.5	6.2	92.3	195	2	2	139	116	45.7
83	7.63	52.1	5.5	11.6	61.1	197	2	4	109	110	45.7
84	8.77	54.5	4.7	5.2	47.0	143	2	4	85	87	25.7
85	8.09	56.9	1.7	7.6	56.9	92	2	3	61	61	45.7
86	9.05	51.2	4.1	20.5	79.8	195	2	3	127	112	45.7
87	7.91	52.8	2.9	11.9	79.5	477	2	3	349	188	65.7
88	10.39	54.6	4.3	14.0	88.3	353	2	2	223	200	65.7
89	9.36	54.1	4.8	18.3	90.6	165	2	1	127	158	45.7
90	11.41	50.4	5.8	23.8	73.0	424	1	3	359	335	45.7
91	8.86	51.3	2.9	9.5	87.5	100	2	3	65	53	25.7
92	8.93	56.0	2.0	6.2	72.5	95	2	3	59	56	25.7
93	8.92	53.9	1.3	2.2	79.5	56	2	2	40	14	5.7
94	8.15	54.9	5.3	12.3	79.8	99	2	4	55	71	25.7
95	9.77	50.2	5.3	15.7	89.7	154	2	2	123	148	25.7
96	8.54	56.1	2.5	27.0	82.5	98	2	1	57	75	45.7
97	8.66	52.8	3.8	6.8	69.5	246	2	3	178	177	45.7
98	12.01	52.8	4.8	10.8	96.9	298	2	1	237	115	45.7
99	7.95	51.8	2.3	4.6	54.9	163	2	3	128	93	42.9
100	10.15	51.9	6.2	16.4	59.2	568	1	3	452	371	62.9
101	9.76	53.2	2.6	6.9	80.1	64	2	4	47	55	22.9
102	9.89	45.2	4.3	11.8	108.7	190	2	1	141	112	42.9
103	7.14	57.6	2.7	13.1	92.6	92	2	4	40	50	22.9
104	13.95	65.9	6.6	15.6	133.5	356	2	1	308	182	62.9
105	9.44	52.5	4.5	10.9	58.5	297	2	3	230	263	42.9
106	10.80	63.9	2.9	1.6	57.4	130	2	3	69	62	22.9
107	7.14	51.7	1.4	4.1	45.7	115	2	3	90	[illegible]	22.9
108	8.02	55.0	2.1	3.8	46.5	91	2	2	44	32	22.9
109	11.80	53.8	5.7	9.1	116.9	571	1	2	441	469	62.9
110	9.50	49.3	5.8	42.0	70.9	98	2	3	68	46	22.9
111	7.70	56.9	4.4	12.2	67.9	129	2	4	85	136	62.9
112	17.94	56.2	5.9	26.4	91.8	835	1	1	791	407	62.9
113	9.41	59.5	3.1	20.6	91.7	29	2	3	20	22	22.9

مجموعة بيانات (ب-٢) *SMSA*

تقدم مجموعة البيانات هذه معلومات حول 141 مساحة حضرية إحصائية قياسية ضخمة في الولايات المتحدة (SMSA). وتشمل المساحة الإحصائية الحضرية القياسية مدينة (أو مدن) بحجم سكاني محدد وتتشكل من مدينة مركزية والمقاطعة (أو المقاطعات) التي تقع فيها المدينة، بالإضافة إلى مقاطعات مجاورة عندما تحقق العلاقات الاجتماعية والاقتصادية بين المقاطعات المركزية والمقاطعات المجاورة معايير محددة من التكامل والمميزات الحضرية. ويمكن أن تتضمن *SMSA* عددا من المدن يصل إلى ثلاثة مدن، كما يمكن أن تعتبر حدود ولاية.

ويتضمن كل سطر من مجموعة البيانات رقم تسلسلي كما يقدم معلومات حول ١١ من المتغيرات الأخرى الخاصة بمساحة بمفردها (SMSA). وتتعلق المعلومات بصورة عامة بالعامين ١٩٧٦ و ١٩٧٧ والمتغيرات الـ ١٢ هي:

رقم المتغير	اسم المتغير	وصف المتغير
١	رقم تسلسل	1 - 141
٢	مساحة الأرض	بالأميال المربعة
٣	عدد السكان	كما هو مقدّر عام ١٩٧٧م (بالآلاف)
٤	النسبة المئوية من السكان في مدن مركزية	النسبة المئوية من سكان الـ SMSA عام ١٩٧٦م في مدينة أو مدن مركزية.
٥	النسبة المئوية لسكان بعمر ٦٥ أو أكبر	النسبة المئوية من سكان الـ SMSA عام ١٩٧٦ ممن أعمارهم ٦٥ سنة فأكثر
٦	عدد الأطباء العاملين	عدد الأطباء غير الاتحاديين الناشطين مهنيا حتى ٣١ ديسمبر ١٩٧٧م.
٧	عدد الأسرة في مستشفى	العدد الكلي للأسِرّة، وأسِرّة الأطفال وأسِرّة الأطفال الشبيهة بالسلّة خلال عام ١٩٧٧م.

رقم المتغير	اسم المتغير	وصف المتغير
٨	النسبة المئوية للمتخرجين من المرحلة الثانوية	النسبة المئوية من السكان البالغين (أشخاص أعمارهم ٢٥ سنة فأكثر) الذين أتموا ١٢ سنة تعليم أو أكثر وفقا لتعداد عام ١٩٧٠ السكاني
٩	القوة العاملة المدنية	العدد الكلي للأشخاص في القوة العاملة المدنية (أشخاص أعمارهم ١٦ سنة فأكثر مصنفون كعاملين أو كعاطلين عن العمل) في ١٩٧٧م (بالآلاف)
١٠	الدخل الشخصي الإجمالي	الدخل الراهن الإجمالي الذي يتلقاه المقيمون في الـ *SMSA* من جميع المصادر عام ١٩٧٦م قبل اقتطاع ضريبة الدخل والضرائب الشخصية للضمان الاجتماعي وبرامج التأمين الاجتماعي الأخرى (بملايين الدولارت)
١١	عدد الجرائم الخطرة	العدد الكلي للجرائم الخطرة في ١٩٧٧، بما في ذلك جرائم القتل، الاغتصاب، السرقة، الاعتداء، السطو على المنازل، اللصوصية وسرقة السيارات، كما أفادت عنها وكالات الأمن.
١٢	منطقة جغرافية	تصنيف المنطقة الجغرافية هو التصنيف المستخدم في مكتب التعداد في الولايات المتحدة، حيث: $NE = 1$، $NC = 2$، $S = 3$ و $W = 4$

المصدر: U.S. Bureau of the Census, *State and Metropolitan Area Data Book*, 1979 (a Statistical Abstract Supplement).

1	2	3	4	5	6	7	8	9	10	11	12
1	1384	9387	78.1	12.3	25627	69678	50.1	4083.9	72100	709234	1
2	4069	7031	44.0	10.0	15389	39699	62.0	3353.6	52737	499813	4
3	3719	7017	43.9	9.4	13326	43292	53.9	3305.9	54542	393162	2
4	3553	4794	37.4	10.7	9724	33731	50.6	2066.3	33216	198102	1
5	3916	4370	29.9	8.8	6402	24167	52.2	1966.7	32906	294466	2
6	2480	3182	31.5	10.5	8502	16751	66.1	1514.5	26573	255162	4
7	2815	3033	23.1	6.7	7340	16941	68.3	1541.9	25663	177355	3
8	1218	2688	0.0	8.8	5255	22137	62.9	1213.3	21524	127567	1
9	8360	2673	46.3	8.2	4047	14347	53.6	1321.2	18350	193125	3
10	6794	2512	60.1	6.3	4562	14333	51.7	1272.7	18221	162976	3
11	4935	2380	21.8	11.0	4071	17752	47.8	1061.2	16120	137479	2
12	3049	2294	19.5	12.1	4005	21149	53.4	967.5	15826	69989	1
13	2259	2147	38.6	9.3	5141	16485	44.6	966.8	14246	138214	3
14	4647	2037	31.5	9.2	3916	12815	65.1	1032.2	14542	112642	2
15	1008	1969	16.6	10.3	4006	16704	55.9	935.5	15953	106646	1
16	1519	1950	31.8	10.5	4094	12545	54.6	906.0	14684	102816	2
17	4326	1832	23.6	7.3	3064	9976	50.4	867.2	12107	106482	3
18	782	1801	28.4	7.8	3119	8656	70.5	915.2	12591	113821	4
19	4261	1683	48.6	9.7	3396	7552	65.3	644.3	10392	112359	4
20	4651	1464	38.8	7.7	3380	8517	67.4	729.2	10375	116861	4
21	2042	1441	24.5	16.5	4071	10039	51.9	681.7	10166	116304	3
22	4226	1427	38.1	9.8	3285	5392	67.8	699.8	10918	91399	4
23	1456	1427	46.7	10.4	2484	8555	56.8	710.4	10104	63695	2
24	2045	1380	37.2	21.4	1949	8863	50.7	543.2	7989	89257	3
25	2149	1375	29.8	10.6	2530	8354	48.4	617.6	9037	68319	2
26	1590	1313	30.1	10.9	2296	9988	50.4	565.7	8411	67965	1
27	27293	1306	25.3	12.3	2018	6323	57.4	510.6	7399	99293	4
28	3341	1293	35.8	10.1	2289	7593	59.9	656.3	9106	81510	2
29	9155	1254	53.8	11.1	2280	6450	60.1	575.2	7766	107370	4
30	1300	1217	47.6	6.8	2794	4989	69.0	610.8	9215	76570	4
31	3072	1144	68.0	9.3	2181	7497	56.0	549.6	7736	61381	2
32	1967	1133	51.1	8.8	2520	8467	45.8	460.5	7030	69285	3
33	3650	1121	34.6	11.1	2358	6224	62.9	539.3	7792	77316	4
34	2460	1087	49.6	8.4	1874	7706	59.9	510.7	6658	62603	2
35	2527	1025	78.7	8.4	1760	7664	46.5	391.1	5582	62694	3
36	2966	970	26.9	10.3	2053	6604	56.3	450.4	6966	54854	1
37	3434	929	28.9	8.3	1844	3215	65.1	422.6	5909	72410	4
38	1392	883	37.2	9.8	1579	6087	46.5	396.8	5705	45642	3
39	2298	886	76.2	9.0	1644	7673	48.2	394.6	5185	52094	3
40	1219	864	31.7	20.6	1396	6158	55.4	352.8	5879	68109	3
41	1708	833	24.0	8.8	1062	5315	56.2	367.5	5489	52606	2
42	8565	822	29.7	7.3	1604	3485	67.6	349.3	4655	49111	4
43	3358	805	35.1	11.3	1649	5512	44.9	359.1	4941	42786	3
44	2624	794	30.4	12.2	1532	4730	55.2	356.5	5094	30771	1
45	2187	777	47.0	10.2	1098	4342	51.9	355.4	5142	46213	2
46	3214	774	47.7	9.4	1285	3459	40.3	401.7	4924	34941	3
47	3491	769	48.5	9.7	1496	5620	59.6	362.3	4798	44513	3
48	4080	773	59.6	9.9	1597	7496	47.3	380.9	4600	33936	3
49	596	723	100.0	6.0	1260	2819	66.0	319.9	5181	46984	4
50	3199	694	80.6	8.7	983	4749	50.8	292.4	4127	43010	3
51	903	661	37.3	9.6	948	4064	55.6	293.3	4102	34725	2
52	2419	647	27.8	9.9	1250	2870	57.8	286.8	3860	30829	1
53	938	644	48.1	7.4	614	3016	50.0	280.9	4177	35106	2
54	1951	629	28.4	14.5	696	4843	47.9	271.5	3667	14868	1
55	1490	624	33.1	11.9	827	3818	47.4	300.2	4144	19090	1
56	5677	610	55.8	10.5	760	3883	56.2	292.0	4035	32146	3
57	1525	597	55.7	8.3	751	3234	44.9	318.5	3777	37070	3
58	2528	593	19.2	10.2	798	3135	55.4	274.1	3489	44442	3
59	312	594	19.5	7.5	769	2463	55.0	298.7	4352	29100	1
60	1537	581	63.8	8.7	1234	5160	62.7	272.6	3725	32271	2
61	1420	576	32.6	9.5	833	2950	54.0	280.8	3553	26645	2
62	47	564	41.9	11.9	745	3352	36.3	258.9	3915	29157	1
63	1023	541	35.1	10.0	639	3144	52.1	234.1	3437	22111	2
64	2115	526	19.9	9.1	676	2296	38.8	253.3	2962	30684	3
65	1182	514	32.4	7.4	518	2515	52.4	216.8	3627	35201	2
66	1165	516	14.5	8.6	746	4277	54.4	237.1	3724	31358	3
67	476	492	8.9	10.9	787	2778	60.1	218.4	3603	24787	1
68	1553	487	50.0	8.0	2207	4931	52.0	257.2	2991	24269	3
69	2023	477	22.1	21.8	752	2317	55.7	194.2	3283	36418	3
70	2766	474	67.9	7.7	679	3873	56.3	224.0	2598	29967	3

1	2	3	4	5	6	7	8	9	10	11	12
71	5966	472	39.5	9.6	737	1907	52.7	246.6	3007	38205	4
72	1863	468	50.4	7.7	674	2989	63.8	194.8	2747	25159	4
73	192	462	60.5	10.8	617	1789	44.1	212.6	3158	27161	1
74	9240	455	67.0	10.3	1123	2347	63.1	183.6	2598	41649	4
75	2277	455	39.5	7.5	512	1788	61.9	221.1	2853	20053	2
76	1630	449	41.9	10.7	724	4395	50.0	198.0	2445	17596	3
77	1617	435	71.0	6.9	518	2031	54.1	197.9	2617	31539	3
78	1057	435	90.7	6.1	479	2551	51.1	163.4	2012	25650	3
79	1624	429	13.4	11.0	832	2938	55.4	207.8	2885	16985	1
80	1676	423	36.6	9.2	505	3297	60.7	156.3	2689	24266	4
81	2818	425	48.5	9.3	540	2694	42.3	172.8	2162	22374	3
82	2866	408	24.9	10.7	427	2864	39.1	169.1	1987	10425	3
83	4883	402	72.4	7.3	873	2236	64.9	185.2	2353	28171	4
84	966	401	24.9	10.6	427	3192	52.2	174.7	2446	15981	2
85	2109	403	41.2	10.3	520	2539	45.2	183.1	2308	16240	3
86	2449	395	68.4	9.6	681	2864	63.2	207.4	2651	25149	2
87	2618	385	31.7	6.1	836	2159	48.0	145.6	1992	25046	3
88	1465	374	30.3	6.8	598	6456	50.6	164.7	2201	26428	3
89	1704	375	52.1	10.5	379	2491	55.6	173.2	2662	18599	2
90	1750	370	49.3	9.7	446	3472	58.2	176.5	2439	16529	2
91	1489	369	58.8	9.5	911	5720	56.5	175.1	2264	26032	3
92	8152	363	22.3	9.1	405	1254	51.7	165.6	2257	28351	4
93	2207	364	57.3	9.7	356	2167	45.5	165.9	2331	19138	3
94	7874	360	44.4	6.9	398	1365	65.2	174.2	2410	33687	4
95	655	364	75.2	6.6	425	3879	51.6	163.0	2088	15623	3
96	1803	362	35.3	10.4	483	2137	53.7	168.9	2666	16405	2
97	2363	356	53.1	10.6	565	2717	49.3	146.4	1996	19212	3
98	1435	352	13.4	11.7	342	1076	44.7	156.8	2165	11273	1
99	946	348	16.4	11.1	366	1455	43.9	163.8	2178	8116	1
100	1136	333	58.6	9.7	448	2630	68.1	171.4	2396	20465	2
101	2658	327	39.0	12.2	365	5430	49.9	136.9	1862	9325	1
102	228	317	31.1	10.2	667	3179	52.8	156.5	2264	19410	1
103	1758	310	56.8	11.5	565	2081	65.3	131.2	1939	17379	4
104	1198	313	55.1	8.0	1171	3877	71.2	172.3	2038	18676	2
105	1412	311	39.2	11.3	436	1837	49.4	154.2	2098	25714	4
106	2071	306	19.9	11.3	470	2531	58.9	133.1	1782	11161	1
107	862	302	26.3	13.4	423	1929	43.3	145.5	2010	7699	1
108	1526	303	71.7	7.7	413	1636	47.1	125.8	1692	20038	3
109	1758	297	33.2	11.6	296	2652	45.3	114.4	1641	12467	3
110	1651	296	64.6	8.9	774	5431	56.1	136.9	1724	14468	3
111	1493	294	64.8	8.9	863	3289	53.7	154.7	1787	15871	3
112	1610	294	59.8	9.5	471	4633	62.9	116.1	1851	18651	4
113	2710	288	63.7	6.2	357	1277	72.8	110.9	1639	18173	4
114	1975	291	46.5	12.6	405	2896	51.5	133.8	1853	12787	2
115	1920	291	49.8	7.8	283	1306	53.2	126.9	1553	12315	3
116	1404	289	38.5	10.0	299	1766	56.2	138.6	1776	11715	2
117	2737	287	45.0	10.5	602	1462	71.3	131.4	1980	18208	4
118	1700	287	18.8	8.0	739	3381	45.9	120.4	1616	14534	3
119	909	277	41.2	11.5	307	1309	54.2	131.9	1762	13722	2
120	1858	277	24.3	13.7	354	1562	46.3	116.9	1507	19133	3
121	3324	275	49.7	8.4	373	929	62.5	120.5	1918	14776	4
122	1697	274	23.8	7.2	338	1610	51.0	105.9	1354	19317	3
123	813	272	46.0	9.8	293	1693	58.4	119.9	1688	10402	1
124	7397	267	47.3	12.5	355	2042	56.2	113.7	1654	12273	2
125	1165	268	43.7	9.4	450	2070	57.5	129.4	1719	16226	2
126	802	268	52.6	9.8	392	1425	52.2	129.6	1816	13230	2
127	1770	268	14.8	12.2	285	2804	44.1	106.7	1537	4205	1
128	495	264	50.7	7.8	220	1177	52.6	119.5	1661	8398	2
129	1255	261	26.0	10.7	458	1646	51.6	113.0	1725	10208	3
130	1148	589	45.3	11.1	891	5790	54.0	277.0	3510	29237	1
131	1509	643	37.6	12.0	1087	4900	51.4	319.6	3982	29058	1
132	2013	254	61.7	9.7	273	1484	50.9	106.7	1412	14446	3
133	711	250	42.4	6.1	1411	3659	67.5	131.0	1790	16228	2
134	471	251	46.3	8.6	219	1128	47.8	105.3	1458	13474	2
135	4552	249	54.4	9.1	329	719	61.9	118.0	1386	15596	4
136	1400	242	50.8	8.0	290	1271	45.7	104.4	1351	10391	3
137	1511	236	38.7	10.7	348	1093	50.4	127.2	1452	16676	4
138	1543	232	39.6	8.1	159	481	30.3	80.6	769	8436	3
139	1011	233	37.8	10.5	264	964	70.7	93.2	1337	14018	3
140	813	232	13.4	10.9	371	4355	58.0	97.0	1589	8428	1
141	654	231	28.8	3.9	140	1296	55.1	66.9	1148	15884	3

تعريف مصادر الـ SMSA

1 NEW YORK, NY
2 LOS ANGELES, CA
3 CHICAGO, IL
4 PHILADELPHIA, PA
5 DETROIT, MI
6 SAN FRANCISCO, CA
7 WASHINGTON, DC
8 NASSAU, NY
9 DALLAS, TX
10 HOUSTON, TX
11 ST. LOUIS, MO
12 PITTSBURG, PA
13 BALTIMORE, MD
14 MINNEAPOLIS, MN
15 NEWARK, NJ
16 CLEVELAND, OH
17 ATLANTA, GA
18 ANAHEIM, CA
19 SAN DIEGO, CA
20 DENVER, CO
21 MIAMI, FL
22 SEATTLE, WA
23 MILWAUKEE, WI
24 TAMPA, FL
25 CINCINNATI, OH
26 BUFFALO, NY
27 RIVERSIDE, CA
28 KANSAS CITY, MO
29 PHOENIX, AZ
30 SAN JOSE, CA
31 INDIANAPOLIS, IN
32 NEW ORLEANS, LA
33 PORTLAND, OR
34 COLUMBUS, OH
35 SAN ANTONIO, TX
36 ROCHESTER, NY
37 SACRAMENTO, CA
38 LOUISVILLE, KY
39 MEMPHIS, TN
40 FT. LAUDERDALE, FL
41 DAYTON, OH
42 SALT LAKE CITY, UT
43 BIRMINGHAM, AL
44 ALBANY, NY
45 TOLEDO, OH
46 GREENSBORO, NC
47 OKLAHOMA CITY, OK
48 NASHVILLE, TN
49 HONOLULU, HI
50 JACKSONVILLE, FL
51 AKRON, OH
52 SYRACUSE, NY
53 GARY, IN
54 NORTHEAST, PA
55 ALLENTOWN, PA
56 TULSA, OK
57 CHARLOTTE, NC
58 ORLANDO, FL
59 NEW BRUNSWICK, NJ
60 OMAHA, NE
61 GRAND RAPIDS, MI
62 JERSEY CITY, NJ
63 YOUNGSTOWN, OH
64 GREENVILLE, SC
65 FLINT, MI
66 WILMINGTON, DE
67 LONG BRANCH, NJ
68 RALEIGH, NC
69 W. PALM BEACH, FL
70 AUSTIN, TX
71 FRESNO, CA
72 OXNARD, CA
73 PATERSON, NJ
74 TUCSON, AZ
75 LANSING, MI
76 KNOXVILLE, TN
77 BATON ROUGE, LA
78 EL PASO, TX
79 HARRISBURG, PA
80 TACOMA, WA
81 MOBILE, AL
82 JOHNSON CITY, TN
83 ALBUQUERQUE, NM
84 CANTON, OH
85 CHATANOOGA, TN
86 WICHITA, KS
87 CHARLESTON, SC
88 COLUMBIA, SC
89 DAVENPORT, IA
90 FORT WAYNE, IN
91 LITTLE ROCK, AR
92 BAKERSFIELD, CA
93 BEAUMONT, TX
94 LAS VEGAS, NV
95 NEWPORT NEWS, VA
96 PEORIA, IL
97 SHREVEPORT, LA
98 YORK, PA
99 LANCASTER, PA
100 DES MOINES, IA
101 UTICA, NY
102 TRENTON, NJ
103 SPOKANE, WA
104 MADISON, WI
105 STOCKTON, CA
106 BINGHAMTON, NY
107 READING, PA
108 CORPUS CHRISTI, TX
109 HUNTINGTON, WV
110 JACKSON, MS
111 LEXINGTON, KY
112 VALLEJO, CA
113 COLORADO SPRINGS, CO
114 EVANSVILLE, IN
115 HUNTSVILLE, AL
116 APPLETON, WI
117 SANTA BARBARA, CA
118 AUGUSTA, GA
119 SOUTH BEND, IN
120 LAKELAND, FL
121 SALINAS, CA
122 PENSACOLA, FL
123 ERIE, PA
124 DULUTH, MN
125 KALAMAZOO, MI
126 ROCKFORD, IL
127 JOHNSTOWN, PA
128 LORAIN, OH
129 CHARLESTON, WV
130 SPRINGFIELD, MA
131 WORCESTER, MA
132 MONTGOMERY, AL
133 ANN ARBOR, MI
134 HAMILTON, OH
135 EUGENE, OR
136 MACON, GA
137 MODESTO, CA
138 MCALLEN, TX
139 MELBOURNE, FL
140 POUGHKEEPSIE, NY
141 FAYETTEVILLE, NC

مجموعات بيانات (ب - ٣) تجربة تأثيرات المخدرات

تقدم هذه المجموعة من البيانات نتائج مأخوذة من تجربة دُرست فيها آثار المخدر على سلوك الفئران. السلوك المدروس هو معدل ضغط فأر محروم من الماء لذراع رافعة كي يحصل على الماء. وقد نُفذت التجربة في جزئين. ويحدد المتغير ٢ جزئي الدراسة (1,2).

في الجزء I من الدراسة استُخدم ١٢ من الفئران البيضاء الذكور من السلالة نفسها ولها تقريبا الوزن نفسه. ويحدد المتغير ٣ كل فأر (1,...,12)، وقبل التجربة دُرِّب كل فأر على ضغط ذراع رافعة للحصول على الماء حتى بلوغ معدل مستقر للضغط. ودُرس في هذه التجربة عاملان - المعدل الابتدائي لضغط الذراع (عامل A) واستخدام المخدر (عامل B). وقد صُنِّفت الفئران الـ ١٢ إلى إحدى ثلاث مجموعات وفقا للمعدل الابتدائي لضغط الذراع (1, 2, 3). المستوى الأول هو معدل بطيء المستوى، الثاني معدل معتدل، والمستوى الثالث معدل سريع، وقد عُرفت المستويات بحيث يصنّف ثلث الفئران إلى كل من المستويات الثلاثة.

ودُرست أربعة مستويات لجرعة المخدر، بما في ذلك المستوى 0 المؤلف من محلول ملحي. ويحدد المتغير ٥ جرعة المخدر (1,...,4). وكل مستويات الجرعة محددة بدلالة الملليغرام من المخدر لكل كيلو غرام من وزن الفأر.

وبعد ساعة من حقن المخدر بدأ دورة تجريبية يتلقى الفأر خلالها الماء في كل مرة بعد الضغط الثاني للذراع. وسنرمز لهذا البرنامج التنفيذي بالرمز FR - 2. وقد تلقى كل فأر المستويات الأربعة من الجرعات بترتيب عشوائي. وقد أعطيت كل جرعة مخدر مرتين لكل فأر مما أتاح وحدتي مشاهدة لكل معالجة. ويحدد المتغير ٦ وحدة المشاهدة (1,2).

وقد عُرف متغير الاستجابة بأنه عدد المرات، الكلي لضغط الذراع مقسوما على الزمن المنصرم بالثواني خلال دورة تجريبية لمعالجة معينة. والمتغير ٧ هو متغير الاستجابة.

وفي الجزء الثاني من الدراسة استُخدم ١٢ فأرا ذكرا أبيض آخر من السلالة نفسها والوزن نفسه تقريبا المستخدم في الجزء I. ويحدد المتغير 2 هذا الجزء من الدراسة والمتغير ٣ يحدد الفئران الـ ١٢ الإضافية (13,...,24). والتصميم التجريبي للجزء II من

الدراسة كان متطابقا بالضبط مع الجزء I، باستثناء أن كل فأر يتلقى الماء في كل مرة بعد الضغط الخامس للذراع. وسنرمز لهذا البرنامج التنفيذي بالرمز $FR - 5$. ويحدد المتغير ٢ البرنامج التنفيذي باعتبار أن الجزء I من الدراسة استخدم البرنامج $FR - 2$ بينما استخدم جزؤها الثاني البرنامج $FR - 5$. وهكذا يشكل البرنامج التنفيذي عاملا آخر (العامل C) تمت دراسته في التجربة المركبة بجزئيها.

ونلخص فيما يلي المتغيرات لهذا التصميم التجريبي:

رقم المتغير	اسم المتغير	وصف المتغير
١	رقم تسلسلي	1 - 192
٢	جزء الدراسة (العامل C : برنامج تنفيذي)	الجزء الأول FR - 2 : 1 الجزء الثاني FR - 2 : 2
٣	هوية الفأر	1 - 24
٤	المعدل الابتدائي لضغط الذراع (عامل A)	بطيء: 1 معتدل: 2 سريع: 3
٥	مستوى الجرعة (مغ/كغ) (عامل B)	محلول ملحي 0 : 1 2 :0.5 3 : 1.0 4 : 1.8
٦	وحدة مشاهدة	1,2
٧	متغير الاستجابة ـ معدل ضغط الذراع (عدد المرات الكلي لضغط الذراع مقسوما على الزمن المنصرم بالثواني)	

المصدر: T.G. Heffner, R.B. Drawbaugh, and M.J. Zigmond, "Amphetamine and Operant Behavior in Rats: Relationship between Drug Effect and Contrl Response Rate," *Journal of Comparative and Physiological Psychology* 86 (1974), pp. 1031-43.

1	2	3	4	5	6	7
1	1	1	1	1	1	.81
2	1	1	1	2	1	.80
3	1	1	1	3	1	.82
4	1	1	1	4	1	.50
5	1	2	1	1	1	.77
6	1	2	1	2	1	.78
7	1	2	1	3	1	.79
8	1	2	1	4	1	.51
9	1	3	1	1	1	.80
10	1	3	1	2	1	.82
11	1	3	1	3	1	.83
12	1	3	1	4	1	.52
13	1	4	1	1	1	.95
14	1	4	1	2	1	.95
15	1	4	1	3	1	.91
16	1	4	1	4	1	.60
17	1	5	2	1	1	1.03
18	1	5	2	2	1	1.13
19	1	5	2	3	1	1.04
20	1	5	2	4	1	.82
21	1	6	2	1	1	.96
22	1	6	2	2	1	.93
23	1	6	2	3	1	1.02
24	1	6	2	4	1	.63
25	1	7	2	1	1	.98
26	1	7	2	2	1	1.00
27	1	7	2	3	1	.98
28	1	7	2	4	1	.74
29	1	8	2	1	1	1.17
30	1	8	2	2	1	1.20
31	1	8	2	3	1	1.18
32	1	8	2	4	1	.91
33	1	9	3	1	1	1.20
34	1	9	3	2	1	1.24
35	1	9	3	3	1	1.27
36	1	9	3	4	1	.96
37	1	10	3	1	1	1.25
38	1	10	3	2	1	1.23
39	1	10	3	3	1	1.30
40	1	10	3	4	1	1.01
41	1	11	3	1	1	1.23
42	1	11	3	2	1	1.20
43	1	11	3	3	1	1.18
44	1	11	3	4	1	.95
45	1	12	3	1	1	1.31
46	1	12	3	2	1	1.42
47	1	12	3	3	1	1.41
48	1	12	3	4	1	1.08
49	1	1	1	1	2	.84
50	1	1	1	2	2	.85
51	1	1	1	3	2	.88
52	1	1	1	4	2	.58
53	1	2	1	1	2	.72
54	1	2	1	2	2	.73
55	1	2	1	3	2	.74
56	1	2	1	4	2	.42
57	1	3	1	1	2	.73
58	1	3	1	2	2	.76
59	1	3	1	3	2	.75
60	1	3	1	4	2	.48
61	1	4	1	1	2	.89
62	1	4	1	2	2	.90
63	1	4	1	3	2	.97
64	1	4	1	4	2	.67
65	1	5	2	1	2	1.11
66	1	5	2	2	2	1.02
67	1	5	2	3	2	1.12
68	1	5	2	4	2	.75
69	1	6	2	1	2	1.01
70	1	6	2	2	2	1.05
71	1	6	2	3	2	.95
72	1	6	2	4	2	.72
73	1	7	2	1	2	1.05
74	1	7	2	2	2	1.07
75	1	7	2	3	2	1.05
76	1	7	2	4	2	.79
77	1	8	2	1	2	1.12
78	1	8	2	2	2	1.13
79	1	8	2	3	2	1.11
80	1	8	2	4	2	.83
81	1	9	3	1	2	1.28
82	1	9	3	2	2	1.17
83	1	9	3	3	2	1.21
84	1	9	3	4	2	.91
85	1	10	3	1	2	1.21
86	1	10	3	2	2	1.31
87	1	10	3	3	2	1.22
88	1	10	3	4	2	.93
89	1	11	3	1	2	1.16
90	1	11	3	2	2	1.15
91	1	11	3	3	2	1.23
92	1	11	3	4	2	1.02
93	1	12	3	1	2	1.40
94	1	12	3	2	2	1.33
95	1	12	3	3	2	1.35
96	1	12	3	4	2	1.20

1	2	3	4	5	6	7
97	2	13	1	1	1	2.18
98	2	13	1	2	1	2.44
99	2	13	1	3	1	1.92
100	2	13	1	4	1	.92
101	2	14	1	1	1	2.02
102	2	14	1	2	1	2.20
103	2	14	1	3	1	1.75
104	2	14	1	4	1	.82
105	2	15	1	1	1	2.06
106	2	15	1	2	1	2.28
107	2	15	1	3	1	1.86
108	2	15	1	4	1	.80
109	2	16	1	1	1	2.28
110	2	16	1	2	1	2.46
111	2	16	1	3	1	1.90
112	2	16	1	4	1	.90
113	2	17	2	1	1	2.62
114	2	17	2	2	1	2.58
115	2	17	2	3	1	2.21
116	2	17	2	4	1	1.03
117	2	18	2	1	1	2.60
118	2	18	2	2	1	2.60
119	2	18	2	3	1	2.34
120	2	18	2	4	1	1.14
121	2	19	2	1	1	2.39
122	2	19	2	2	1	2.41
123	2	19	2	3	1	2.09
124	2	19	2	4	1	.90
125	2	20	2	1	1	2.70
126	2	20	2	2	1	2.64
127	2	20	2	3	1	2.23
128	2	20	2	4	1	1.02
129	2	21	3	1	1	2.98
130	2	21	3	2	1	2.64
131	2	21	3	3	1	2.34
132	2	21	3	4	1	1.28
133	2	22	3	1	1	3.10
134	2	22	3	2	1	2.85
135	2	22	3	3	1	2.40
136	2	22	3	4	1	1.35
137	2	23	3	1	1	2.80
138	2	23	3	2	1	2.48
139	2	23	3	3	1	2.16
140	2	23	3	4	1	1.01
141	2	24	3	1	1	3.21
142	2	24	3	2	1	2.92
143	2	24	3	3	1	2.56
144	2	24	3	4	1	1.40
145	2	13	1	1	2	2.26
146	2	13	1	2	2	2.40
147	2	13	1	3	2	1.99
148	2	13	1	4	2	.99
149	2	14	1	1	2	1.96
150	2	14	1	2	2	2.18
151	2	14	1	3	2	1.81
152	2	14	1	4	2	.78
153	2	15	1	1	2	2.10
154	2	15	1	2	2	2.24
155	2	15	1	3	2	1.92
156	2	15	1	4	2	.88
157	2	16	1	1	2	2.35
158	2	16	1	2	2	2.49
159	2	16	1	3	2	1.95
160	2	16	1	4	2	.96
161	2	17	2	1	2	2.68
162	2	17	2	2	2	2.64
163	2	17	2	3	2	2.17
164	2	17	2	4	2	.96
165	2	18	2	1	2	2.66
166	2	18	2	2	2	2.62
167	2	18	2	3	2	2.28
168	2	18	2	4	2	1.23
169	2	19	2	1	2	2.43
170	2	19	2	2	2	2.48
171	2	19	2	3	2	2.16
172	2	19	2	4	2	.84
173	2	20	2	1	2	2.66
174	2	20	2	2	2	2.70
175	2	20	2	3	2	2.27
176	2	20	2	4	2	.98
177	2	21	3	1	2	2.94
178	2	21	3	2	2	2.70
179	2	21	3	3	2	2.44
180	2	21	3	4	2	1.33
181	2	22	3	1	2	3.20
182	2	22	3	2	2	2.91
183	2	22	3	3	2	2.45
184	2	22	3	4	2	1.39
185	2	23	3	1	2	2.84
186	2	23	3	2	2	2.53
187	2	23	3	3	2	2.23
188	2	23	3	4	2	1.07
189	2	24	3	1	2	3.31
190	2	24	3	2	2	2.98
191	2	24	3	3	2	2.47
192	2	24	3	4	2	1.51

الملحق جـ

مختارات من المراجع

المراجع المختارة مصنفة في خمسة أصناف :

١ ـ كتب انحدار عامة.

٢ ـ تشخيصات وبناء نماذج.

٣ ـ حسابات إحصائية.

٤ ـ كتب عامة في التصميم التجريبي وتحليل التباين.

٥ ـ مواضيع متفرقة.

١ ـ كتب انحدار عامة

llen, D. M., and F. B. Cady. *Analyzing Experimental Data by Regression*. New York: Van Nostrand Reinhold, 1982.

owerman, B. L.; R. T. O'Connell; and D. A. Dickey. *Linear Statistical Models: An Applied Approach*. Boston: Duxbury Press, 1986.

rook, R. J., and G. C. Arnold. *Applied Regression Analysis and Experimental Design*. New York: Marcel Dekker, 1985.

.hatterjee, S., and B. Price. *Regression Analysis by Example*. New York: John Wiley & Sons, 1977.

:ohen, J., and P. Cohen. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983.

aniel, C., and F. S. Wood. *Fitting Equations to Data*. 2nd ed. New York: Wiley-Interscience, 1980.

raper, N.R., and H. Smith. *Applied Regression Analysis*. 2nd ed. New York: John Wiley & Sons, 1981.

unn, O. J., and V. A. Clark. *Applied Statistics: Analysis of Variance and Regression*. 2nd ed. New York: John Wiley & Sons, 1987.

dwards, A. L. *An Introduction to Linear Regression and Correlation*. 2nd ed. New York: W. H. Freeman & Co., 1984.

Edwards, A. L. *Multiple Regression and the Analysis of Variance and Covariance*. 2nd ed. New York: W. H. Freeman & Co., 1985.

Gunst, R. F., and R. L. Mason. *Regression Analysis and Its Application*. New York: Marcel Dekker, 1980.

Kleinbaum, D. G.; L. L. Kupper; and K. E. Muller. *Applied Regression Analysis and Other Multivariate Methods*. 2nd ed. Boston: PWS-Kent Publishing Co., 1988.

Mendenhall, W., and T. Sincich. *A Second Course in Business Statistics: Regression Analysis*. 2nd ed. San Francisco: Dellen Publishing Co., 1986.

Montgomery, D. C., and E. A. Peck. *Introduction to Linear Regression Analysis*. New York: John Wiley & Sons, 1982.

Mosteller, F., and J. W. Tukey. *Data Analysis and Regression*. Reading, Pa.: Addison-Wesley Publishing, 1977.

Myers, R. H. *Classical and Modern Regression with Applications*. Boston: Duxbury Press, 1986.

Pedhazur, E. J. *Multiple Regression in Behavioral Research*. 2nd ed. New York: Holt, Rinehart & Winston, 1982.

Seber, G. A. F. *Linear Regression Analysis*. New York: John Wiley & Sons, 1977.

Weisberg, S. *Applied Linear Regression*. 2nd ed. New York: John Wiley & Sons, 1985.

Younger, M. S. *A First Course in Linear Regression*. 2nd ed. Boston: Duxbury Press, 1985.

٢ ـ تشخيصات وبناء نماذج

Allen, D. M. "Mean Square Error of Prediction as a Criterion for Selecting Variables." *Technometrics* 13 (1971), pp. 469–75.

Anscombe, F. J., and J. W. Tukey. "The Examination and Analysis of Residuals." *Technometrics* 5 (1963), pp. 141–60.

Atkinson, A. C. *Plots, Transformations and Regression*. Oxford: Clarendon Press, 1985.

Barnett, V., and T. Lewis. *Outliers in Statistical Data*. 2nd ed. New York: John Wiley & Sons, 1984.

Belsley, D. A.; E. Kuh; and R. E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons, 1980.

Box, G. E. P., and D. R. Cox. "An Analysis of Transformations." *Journal of the Royal Statistical Society B* 26 (1964), pp. 211–43.

Box, G. E. P., and N. R. Draper. *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons, 1987.

Box, G. E. P., and P. W. Tidwell. "Transformations of the Independent Variables." *Technometrics* 4 (1962), pp. 531–50.

Chatterjee, S., and A. S. Hadi. *Sensitivity Analysis in Linear Regression*. New York: John Wiley & Sons, 1988.

Cook, R. D., and S. Weisberg. *Residuals and Influence in Regression*. London: Chapman and Hall, 1982.

Durbin, J., and G. S. Watson. "Testing for Serial Correlation in Least Squares Regression. II." *Biometrika* 38 (1951), pp. 159–78.

Flack, V. F., and P. C. Chang. "Frequency of Selecting Noise Variables in Subset-Regression Analysis: A Simulation Study." *The American Statistician* 41 (1987), pp. 84–86.

Freedman, D. A. "A Note on Screening Regression Equations." *The American Statistician* 37 (1983), pp. 152–55.

Glaser, R. E. "Bartlett's Test of Homogeneity of Variances." In *Encyclopedia of Statistical Sciences*, vol. 1, ed. S. Kotz and N. L. Johnson. New York: John Wiley & Sons, 1982, pp. 189–91.

Hoaglin, D. C., and R. Welsch. "The Hat Matrix in Regression and ANOVA." *The American Statistician* 32 (1978), pp. 17–22.

Hocking, R. R. "The Analysis and Selection of Variables in Linear Regression." *Biometrics* 32 (1976), pp. 1–49.

Hoerl, A. E., and R. W. Kennard. "Ridge Regression: Applications to Nonorthogonal Problems." *Technometrics* 12 (1970), pp. 69–82.

Mallows, C. L. "Some Comments on C_p." *Technometrics* 15 (1973), pp. 661–75.

Mansfield, E. R., and M. D. Conerly. "Diagnostic Value of Residual and Partial Residual Plots." *The American Statistician* 41 (1987), pp. 107–16.

Mantel, N. "Why Stepdown Procedures in Variable Selection." *Technometrics* 12 (1970), pp. 621–25.

Pope, P. T., and J. T. Webster. "The Use of an *F*-Statistic in Stepwise Regression Procedures." *Technometrics* 14 (1972), pp. 327–40.

Rousseeuw, P. J., and A. M. Leroy. *Robust Regression and Outlier Detection*. New York: John Wiley & Sons, 1987.

Snee, R. D. "Validation of Regression Models: Methods and Examples." *Technometrics* 19 (1977), pp. 415–28.

Stone, M. "Cross-Validatory Choice and Assessment of Statistical Prediction." *Journal of the Royal Statistical Society B* 36 (1974), pp. 111–47.

Theil, H., and A. L. Nagar. "Testing the Independence of Regression Disturbances." *Journal of the American Statistical Association* 56 (1961), pp. 793–806.

٣ ـ حسابات إحصائية

Dixon, W. J., chief editor. *BMDP Statistical Software Manual*, vols. 1 and 2. Berkeley, Calif.: University of California Press, 1988.

IMSL, Inc. *STAT/LIBRARY User's Manual, Version 1.1*. Houston: IMSL, 1989.

Kennedy, W. J., Jr., and J. E. Gentle. *Statistical Computing*. New York: Marcel Dekker, 1980.

MINITAB Reference Manual, Release 7. State College, Pa.: Minitab, Inc., 1989.

NAG, *The Generalized Linear Interactive Modelling (GLIM) System, Release 3.77*. Downers Grove, Ill.: Numerical Algorithms Group, Inc., 1986.

SAS User's Guide: Statistics. Version 6 ed. Cary, N.C.: SAS Institute, 1987.

SPSS^X User's Guide. 2nd ed. Chicago: SPSS, 1986.

٤ ـ كتب عامة في التصميم التجريبي وتحليل التباين

Anderson, V. L., and R. A. McLean. *Design of Experiments*. New York: Marcel Dekker, Inc., 1974.

Box, G. E. P.; W. G. Hunter; and J. S. Hunter. *Statistics for Experimenters*. New York: John Wiley & Sons, 1978.

Cochran, W. G., and G. M. Cox. *Experimental Designs*. 2nd ed. New York: John Wiley & Sons, 1957.

Cox, D. R. *Planning of Experiments*. New York: John Wiley & Sons, 1958.

Fisher, R. A. *The Design of Experiments*. 8th ed. New York: Hafner Publishing Co., 1966.

Gill, J. L. *Design and Analysis of Experiments*, vols. I and II. Ames, Iowa: Iowa State University Press, 1978.

Graybill, F. A. *Theory and Application of the Linear Model*. Boston: Duxbury Press, 1976.

Hicks, C. R. *Fundamental Concepts in the Design of Experiments*. 3rd ed. New York: Holt, Rinehart and Winston, 1982.

Hocking, R. R. *The Analysis of Linear Models*. Monterey, Calif.: Brooks/Cole Publishing Co., 1985.

John, P. W. M. *Statistical Design and Analysis of Experiments*. New York: Macmillan Co., 1971.

Johnson, N. L., and F. C. Leone. *Statistics and Experimental Design in Engineering and the Physical Sciences*, vols. I and II. 2nd ed. New York: John Wiley & Sons, 1966.

Kempthorne, O. *The Design and Analysis of Experiments*. New York: John Wiley & Sons, 1952.

Keppel, G. *Design and Analysis: A Researcher's Handbook*. 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, 1982.

Kirk, R. E. *Experimental Design: Procedures for the Behavioral Sciences*. 2nd ed. Monterey, Calif.: Brooks/Cole Publishing Co., 1982.

Mendenhall, W. *Introduction to Linear Models and the Design and Analysis of Experiments*. Boston: Duxbury Press, 1968.

Montgomery, D. C. *Design and Analysis of Experiments*. 2nd ed. New York: John Wiley & Sons, 1983.

Myers, J. L. *Fundamentals of Experimental Design*. 3rd ed. Boston: Allyn and Bacon, Inc., 1979.

Peterson, R. G. *Design and Analysis of Experiments*. New York: Marcel Dekker, Inc., 1985.

Scheffé, H. *The Analysis of Variance*. New York: John Wiley & Sons, 1959.

Searle, S. R. *Linear Models for Unbalanced Data*. New York: John Wiley & Sons, 1987.

Seber, G. A. F. *The Linear Hypothesis*. 2nd ed. London: Charles Griffin, 1980.

Steel, R. G. D., and J. H. Torrie. *Principles and Procedures of Statistics*. 2nd ed. New York: McGraw-Hill Book Co., 1980.

Winer, B. J. *Statistical Principles in Experimental Design*. 2nd ed. New York: McGraw-Hill Book Co., 1971.

٥ - مواضيع متفرقة

Berkson, J. "Are There Two Regressions?" *Journal of the American Statistical Association* 45 (1950), pp. 164–80.

Bishop, Y. M. M.; S. E. Fienberg; and P. W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass.: MIT Press, 1975.

Box, G. E. P. "Use and Abuse of Regression." *Technometrics* 8 (1966), pp. 625–29.

Box, G. E. P., and G. M. Jenkins. *Times Series Analysis: Forecasting and Control*. Rev. ed. San Francisco: Holden-Day, 1976.

Cox, D. R. "Notes on Some Aspects of Regression Analysis." *Journal of the Royal Statistical Society A* 131 (1968), pp. 265–79.

Federer, W. T., and M. Zelen. "Analysis of Multifactor Classifications with Unequal Numbers of Observations." *Biometrics* 22 (1966), pp. 525–52.

Fuller, W. A. *Measurement Error Models*. New York: John Wiley & Sons, 1987.

Gibbons, J. D. *Nonparametric Methods for Quantitative Analysis*. 2nd ed. Columbus, Ohio: American Sciences Press, 1985.

Graybill, F. A. *Matrices with Applications in Statistics*. 2nd ed. Belmont, Calif.: Wadsworth, 1983.

Greenhouse, S. W., and S. Geisser. "On Methods in the Analysis of Profile Data." *Psychometrika* 24 (1959), pp. 95–112.

Hocking, R. R. "A Discussion of the Two-Way Mixed Model." *The American Statistician* 27 (1973), pp. 148–52.

Hogg, R. V. "Statistical Robustness: One View of Its Use in Applications Today." *The American Statistician* 33 (1979), pp. 108–15.

Huynh, H., and L. Feldt. "Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split-plot Designs." *Journal of Educational Statistics* 1 (1976), pp. 69–82.

Johnson, D. E., and F. A. Graybill. "Estimation of σ^2 in a Two-Way Classification Model with Interaction." *Journal of the American Statistical Association* 67 (1972), pp. 388–94.

Johnson, R. A., and D. W. Wichern. *Applied Multivariate Statistical Analysis*. 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, 1988.

Koch, G. G.; J. D. Elashoff; and I. A. Amara. "Repeated Measurements—Design and Analysis." In *Encyclopedia of Statistical Sciences*, vol. 8, ed. S. Kotz and N. L. Johnson. New York: John Wiley & Sons, 1988, pp. 46–73.

Miller, R. G., Jr. *Simultaneous Statistical Inference*. 2nd ed. New York: Springer-Verlag, 1981.

Owen, D. B. *Handbook of Statistical Tables*. Reading, Mass.: Addison-Wesley Publishing, 1962.

Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models and Economic Forecasts*. 2nd ed. New York: McGraw-Hill, 1981.

Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2 (1946), pp. 110–14.

Searle, S. R. *Matrix Algebra Useful for Statistics*. New York: John Wiley & Sons, 1982.

Snedecor, G. W., and W. G. Cochran. *Statistical Methods*. 7th ed. Ames, Iowa: Iowa State University Press, 1980.

ثبت المصطلحات

- عربي – إنجليزي
- إنجليزي – عربي

أولا: عربي - إنجليزي

أ

أثر الحافة	dge trace
احتمال شرطي	nditional probability
اختبار إف جزئي	tial f-test
جودة التوفيق	odness of fit test
خطي عام	neral linear test
نقص التوفيق	ck of fit test
ارتباط	rrelation
تسلسلي	ial correlation
جزئي	tial correlation
ذاتي	tocorrelation
متعدد	ltiple correlation
استجابة	sponse
استقلال خطي	ear independence
انحدار	gression
خطي قطعة منقطعة	cewise linear regression
عبر نقطة الأصل	gression through origin
متعدد	ltiple regression
الانحرافات المطلقة الدنيا	st absolute deviation
انحراف كلي	al deviation

ب

بيانات (معطيات)	Data
تجريبية	Experimental data
مشاهدة	Observational data

ت

تأثيرات تجميعية	Additive effects
تباين	Variance
تحقق متبادل من صحة نموذج مؤسس على أحد جزئي البيانات باستخدام الجزء الآخر	Double cross - validation
من صحة نموذج انحدار	Validation of regression model
تحليل التباين	Analysis of variance
تغاير	Analysis of covariance
الراسب	Residual analysis
تحويل ارتباط	Correlation transformation
بوكس - كوكس	Box - Cox transformation
القوة	Power transformation
المتغيرات	Variables transformation
ترميز	Coding
تصميم تام العشوائية	Completely randomized design
تغاير	Covariance
تفاعل	Interaction
تقدير مشترك	Joint estimate
منحاز	Biased estimate

Data splitting	تقسيم البيانات
Replication	تكرار
Ferecasting	تنبؤ
Prediction of new observation	بمشاهدة جديدة
F - distribution	توزيع إف
T - distribution	T
Fitting data	توفيق بيانات
Expectation	توقع

جـ

ANOVA table	جدول تحاين
All possible regressibns	جميع الانحدارات الممكنة

حـ

Complementary event	حادثة متممة
Case	حالة (مشاهدة)
Influential observation	نافذة
Disturbance term	حد شغب
Backward elemination	حذف إلى الوراء
Statistical package	حزمة إحصائية

خـ

Homoseedastisity	خاصية التجانس
Heteroseedastisity	عدم التجانس (التفاوت)
Error	خطأ
Pure error	بحت
Observation error	قياس

Multicollinearity	خطية متعددة
"Best" subset algorithm	خوارزمية المجموعات الجزئية "الأفضل"

د

Joint probability function	دالة احتمال مشتركة
Marginal probability function	هامشية
Quadratic response function	استجابة تربيعية
Likelihood function	إمكانية
Regression function	انحدار
Degree of freedom	درجة حرية

ر

Residual	راسب
Deleted residual	محذوف
Studentized deleted residual	معيّر تقديرا
Standardized residual	معياري
Studentized residual	معيّر تقديرا
Externally studentized residual	بصورة خارجية
Internally studentized residual	داخلية
Rank of a matrix	رتبة مصفوفة
Normal probaility plot	رسم احتمال طبيعي
Seatter plot	انتشار
Partical regression plot	انحدار جزئي
Stem and leaf plot	جذع وورقة
Residual plot	راسب
Partial residual plot	جزئي

ıe plot	زمني
: plot	صندوقي
ts	رسوم (رسومات)

س

ponse surface	سطح استجابة

ص

ıdratic form	صيغة تربيعية

ط

mptotic normality	طبيعية مقاربة
ward selection procedure	طريقة اختيار إلى الأمام
ıwise regression selection procedure	انحدار خطوة فخطوة
ıimum absolute deviation method	الانحرافات المطلقة الصغرى
ferroni joint estimation procedure	بونفيروني للتقدير المشترك
t differences procedure	الفروق الأولى
imum L1 - norm method	الناظم ال واحد الأصغري

ع

ily of estimates	عائلة من التقديرات
ance inflation factor	عامل تضخم التباين
ntitative factor	كمّي
litative factor	وصفي (كيفي)
ar	عدد سلّمي
ɜrage	عزم
ıstical relation	علاقة إحصائية
ːtional relation	دالّية

غ

Non - central	غير مركزي

ف

Prediction interval	فترة تنبؤ
Confidence interval	ثقة
Hyper - plane	فوق المستوى

ق

Fitted value	قيمة توفيقية

ك

Chi-square	مربع كاي
Orthogonal polynomials	كثيرات الحدود المتعامدة

م

Product operator	مؤثر جداء
Row vector	متجه سطر
Column vector	عمود
Instrumental variable	متغير أداة
Response variable	استجابة
Dependent variable	تابع
Prediction variable	تنبؤ
Latent prediction variable	مستتر
Binary variable	ثنائي
Dummy variable	دُمية
Standardized normal variable	طبيعي معياري
Indicator variable	مؤشر
Independent variable	مستقل

.ean squares	متوسط مربعات
egression mean squares	الانحدار
esidual mean squares	الراسب
:ope of a model	مجال نموذج
ım of squares	مجموع مربعات
:tra sum of squares	إضافي
otal sum of squares	كلي
'ial	محاولة ـ تجربة
eterminant of a matrix	محدد مصفوفة
ontour diagrams	مخططات التساوي
:atter diagram	مخطط انتشار
east squares	مربعات دنيا
eighted least squares	مرجحة
:raleizGed least squares	معممة
ıi-square	مربع كاي
ook's distance	مسافة كوك
ired observations	مشاهدات مزدوجة
ıtlier observation	مشاهدة قاصية
.trix	مصفوفة
ıt matrix	القبّعة
.riance - covariance matrix	تباين ـ تغاير
ıgular matrix	شاذة
empotent matrix	متساوية القوى
mmetric matrix	متناظرة
agonal matrix	نظرية
ormal equations	معادلات ناظمية
eatment	معالجة

Correlation coefficient	معامل ارتباط
Regression coefficient	انحدار
Standardized regression coefficient	معياري
Beta coefficient	بيتا
Family confidence coefficient	ثقة عائلي
Standardization	معايرة
PRESS criterion	معيار بريس
Unbiased estimator	مقدِّر غير منحاز
Sufficient estimator	كاف
Biased estimator	منحاز
Consistent estimator	منسق
Point estimator	نقطي
Regression curve	منحنى الانحدار
Transpose of a matrix	منقول مصفوفة

ن

Central limit theorem	نظرية النهاية المركزية
Lack of fit	نقص التوفيق
Linear regression models	نماذج انحدار خطي
Analysis of variance models	تحليل تباين
Covariance models	تغاير
Burgson model	نموذج برجسون
Full model	تام
Autoregressive error model	خطأ ذاتي الانحدار
Linear model	خطي
Reduced model	مخفض

و

Experimental unit	وحدة تجريبية

ثانيا: إنجليزي - عربي

A

Addetive effects	تأثيرات تجميعية
All possiblee regressions	جميع الانحدارات الممكنة
Analysis of covariance	تحليل تغاير
variance	تحليل تباين
variance models	نماذج تحليل تباين
ANOVA table	جدول تحاين
Asymptotic normality	طبيعيّة مقاربة (تقاربية)
Auto correlation	ارتباط ذاتي
regressive error model	نموذج خطأ ذاتي الانحدار

B

Backward elimination	حذف إلى الوراء
Berkson model	نموذج برجسون
"Best" subset algorithm	خوارزمية المجموعات الجزئية "الأفضل"
Beta coefficient	معامل بيتا
Biased estimation	تقدير منحاز
estimator	مقدِّر منحاز
Binary variable	متغير ثنائي
Bonferroni joint estimotion procedure	طريقة بونفيروني للتقدير المشترك
Box - Cox transformation	تحويل بوكس - كوكس
Box plot	رسم صندوقي

C

Calibration	معايرة
Case	حالة ـ مشاهدة

Central limit theorem	نظرية النهائية المركزية
Chi-square	مربع كاي
Coding	ترميز
Coefficient of simple correlation	معامل ارتباط بسيط
Column vector	متجه عمود
Complementary event	حادثة متممة
Completely randomized design	تصميم تام العشوائية
Conditional probability	احتمال شرطي
Confidence interval	فترة نشطة
Consistent estimator	مقدّر منسق
Contour diagrams	مخططات التساوي
Cook's distance	مسافة كوك
Correlation	ارتباط
transformation	تحويل ارتباط
Covariance	تغاير
models	نماذج تغاير

D

Data	بيانات ـ معطيات
splitting	تقسيم البيانات
Degree of freedom	درجة حرية
Deleted residual	راسب محذوف
Dependent variable	متغير تابع
Determinant of a matrix	محدد مصفوفة
Diagonal matrix	مصفوفة قطرية
Disturbance term	حدّ شغب

English	العربية
Double cross - validation	تحقق متبادل من صحة نموذج مؤسس على أحد جزئي البيانات باستخدام الجزء الآخر
Dummy variable	متغير دمية
E	
Error	خطأ
Expectation	توقع
Experimental data	بيانات تجريبية
unit	وحدة تجريبية
Externally studentized residual	راسب معيّر تقديرا بصورة خارجية
Extra sum of squares	مجموع مربعات إضافي
F	
Family confidence coefficient	معامل ثقة عائلي
of estimates	عائلة من التقديرات
F - distribution	توزيع إنّ
First differences procedure	طريقة الفروق الأولى
order regression model	نموذج انحدار عن المرتبة الأولى
Fitted value	قيمة توفيقية
Fitting data	توفيق بيانات
Forecasting	تنبؤ
Forward selection procedure	طريقة الاختيار إلى الأمام
F - test	اختبار إنّ
Full model	نموذج تام
Functional relation	علاقة دالّية
G	
Generalized least square	مربعات دنيا معممة

Generalized linear test	الاختبار الخطي العام
Goodness of fit test	اختبار جودة التوفيق

H

Hat matrix	مصفوفة القبّعة
Heteroscedastisity	خاصية عدم التجانس (التفاوت)
Homoscedastisity	خاصية التجانس
Hyper - plane	فوق المستوى

I

Idempotent	مصفوفة متساوية القوى
Independent variable	متغير مستقل
Indicator variable	متغيرمؤشر
Influential case	حالة نافذة (ذات نفوذ)
Instrumental variable	متغير أداة
Interaction	تفاعل
Internally studentized residual	راسب معير تقديرا بصورة داخلية

J

Joint estimate	تقدير مشترك
probability function	دالة احتمال مشتركة

L

Lack of fit	نقص التوفيق
test	اختبار نقص التوفيق
Latent predictor variable	متغير تنبؤ مستتر
Least absolute deviations	الانحرافات المطلقة الدنيا
squares	مربعات دنيا
Leverage	عزم
Likelihood function	دالة إمكانية

Linear	خطي
Linear independence	استقلال خطي
model	نموذج خطي
regression model	نموذج انحدار خطي

M

Marginal probability function	دالة احتمال هامشية
Matrix	مصفوفة
Mean response	متوسط استجابة
square	متوسط مربعات
Measurement error	خطأ قياس
Minimum absolute deviation method	طريقة الانحرافات المطلقة الصغرى
L1-norm method	طريقة الناظم ال واحد الأصغري
Multicollinearty	خطية متعددة
Multiple correlation	ارتباط متعدد
regression	انحدار متعدد

N

Noncentral	غير مركزي
Normal equations	معادلات ناظمية
probability plot	رسم احتمال طبيعي

O

Observational data	بيانات مشاهدة
Orthogonal polynomials	كثيرات الحدود المتعامدة
Outlier observation	مشاهدة قاصية

P

Paired observations	مشاهدات مزدوجة
Partial correlation	ارتباط جزئي

Partial F - test	اختبار إف جزئي
regression plot	رسم انحدار جزئي
residual plot	رسم راسب جزئي
Piecewise linear regression	انحدار خطي قطعة فقطعة
Plots	رسوم (رسومات)
Point estimator	متعدِّد نقطي
Power transformation	تحويل القوة
Prediction interval	فترة تنبؤ
of new observation	تنبؤ بمشاهدة جديدة
Predictor variable	متغير تنبؤ
Press criterion	معيار بريس
Product operator	مؤثر جداء
Pure error	خطأ بحت

Q

Quadratic form	صيغة تربيعية
response function	دالة استجابة تربيعية
Qualitative factor	عامل وصفي
Quantitative factor	عامل كمي

R

Random matrix	مصفوفة عشوائية
Rank of a matrix	رتبة مصفوفة
Reduced model	نموذج مخفض
Regression	انحدار
coefficient	معامل انحدار
curve	منحني الانحدار
function	دالة انحدار

mean square	متوسط مربعات لانحدار
model	نموذج انحدار
through origin	انحدار عبر نقطة الأصل (المبدأ)
Replication	تكرار
Residual	راسب
analysis	تحليل راسب
mean square	متوسط مربعات الراسب
plot	رسم راسب
Response	استجابة
surface	سطح استجابة
variable	متغيراستجابة
Restricted model	نموذج مقيّد (مخفض)
Ridge regression	انحدار الحافة
trace	أثر الحافة
Row vector	متجه سطر

S

Scalar	عدد سلّمي
Scatter diagram	مخطط انتشار
plot	رسم انتشار
Scope of a model	مجال نموذج
Serial correlation	ارتباط تسلسلي
Singular matrix	مصفوفة شاذة
Standardized regression	معامل انحدار معياري (قياسي)
residual	راسب معياري
Standard normal variable	متغير طبيعي معياري
Statistical package	حزمة إحصائية
relation	علاقة إحصائية

Stem and leaf plot	رسم جذع وورقة
Stepwise regression selection procedure	طريقة اختيار انحدار خطوة فخطوة
Studentized deleted residual	راسب محذوف معيّر تقديرا
residual	راسب معير تقديرا
Sufficient estimator	مقدِّر كافٍ
Sum of squares	مجموع مربعات
Symmetrix matrix	مصفوفة متناظرة

T

T-distribution	توزيع t
Time plot	رسم زمني
Total deviation	انحراف كلي
sum of squares	مجموع مربعات كلي
Transformation of variables	تحويل المتغيرات
Transpose of a matrix	منقول مصفوفة
Treatment	معالجة
Trial	محاولة - تجربة

U

Unbiased estimator	مقدِّر غير منحاز

V

Validation of regression model	تحقق من صحة نموذج انحدار
Variance	تباين
- covariance matrix	مصفوفة تباين - تغاير
inflation factor	عامل تضخّم التباين
Vector	متجه

W

Weighted least squares	مربعات دنيا مرجّحة

كشاف الموضوعات

ن

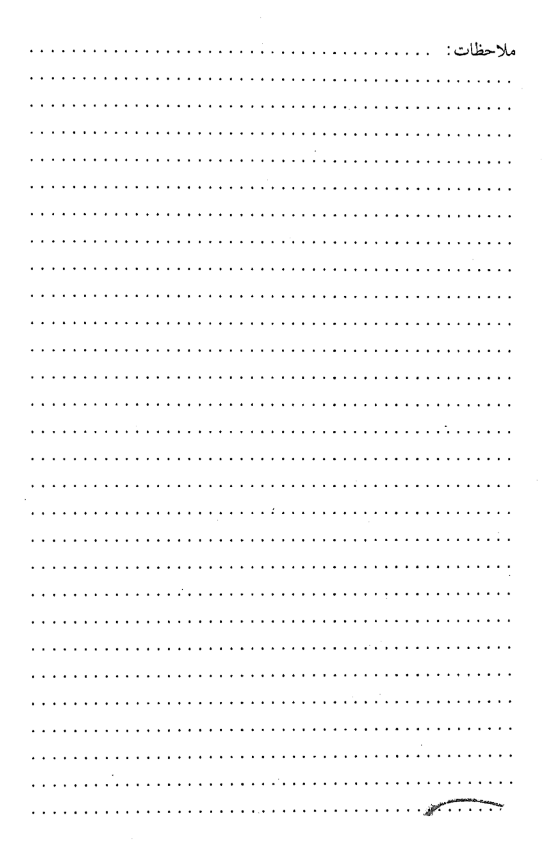


ملاحظات:

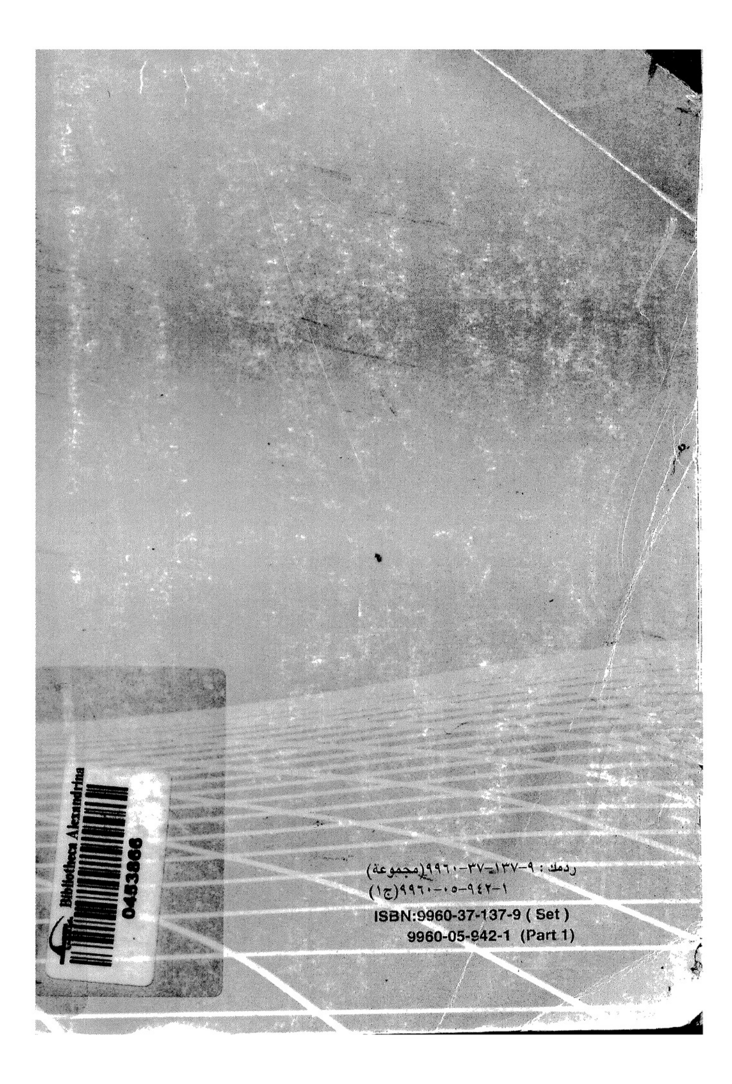
ملاحظات: ....

ملاحظات:

ملاحظات:






Bibliotheca Alexandrina
0453866
ردمك : ٩٩٦٠-٣٧-١٣٧-٩ (مجموعة)
٩٩٦٠-٠٥-٩٤٢-١ (ج١)
ISBN:9960-37-137-9 (Set)
9960-05-942-1 (Part 1)